REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-06-2011	Final Performance Report	20090930 - 20110331

4. TITLE AND SUBTITLE

Quantum System Identification via L1-norm Minimization

5a. CONTRACT NUMBER

5b. GRANT NUMBER

FA9550-09-1-0710

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Robert Kosut, Hersch Rabitz

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SC Solutions, Inc. 1261 Oakmead Pkwy, Sunnyvale CA 94085
Princeton University, 4 New South Bldg, Princeton NJ 08544

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
875 No. Randolph St. RM 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-OSR-VA-TR-2012-0740

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report summarizes our efforts to apply the theory and algorithms of Compressed Sensing (CS) to Quantum Process Tomography (QPT) and Hamiltonian parameter estimation. Specific results include: (1) Development of computational algorithms to include physics based constraints on the quantum process matrix, i.e., positive-semidefinite and trace preserving. (2) Simulations of two-qubit Quantum Fourier Transform interacting with an unknown environment. (3) Establishment of robustness of ideal unitary basis via singular-value-decomposition. (4) The first experimental demonstration of QPT via CS on a photonic system at the University of Queensland. The latter experimental results showed the anticipated and predicted significant reduction of estimation resources, e.g., with respect to an estimate of a 16x16 process matrix obtained from an over complete set of 576 configurations, only 32 configurations were needed to obtain a 97% fidelity, and only 18 configurations to obtain a 94% fidelity. (5) Application of CS to a nearly-sparse many-body Hamiltonian.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Robert Kosut
UNC	UNC	UNC		35	19b. TELEPHONE NUMBER *(Include area code)* 415-601-6287

Final Report:

Quantum System Identification via ℓ_1-norm Minimization

Supported by:
DARPA Grant FA9550-09-1-0710 administered through AFOSR
(30 Sept 2009 – 31 March 2011)

Submitted by:
SC Solutions
1261 Oakmead Parkway, Sunnyvale, CA 94085

Authors:
Robert Kosut (SC Solutions) & Hersch Rabitz (Princeton University)

Technical POC:
Dr. Robert Kosut
415-601-6287, kosut@scsolutions.com

Administrative POC:
Mr. Eric Denham
408-617-4545, eric@scsolutions.com

30 June, 2011

Contents

1 Motivation

> *"I would like to describe a field, in which little has been done, but in which an enormous amount can be done in principle. . . . Furthermore, a point that is most important is that it would have an enormous number of technical applications. What I want to talk about is the problem of manipulating and controlling things on a small scale."* Richard P. Feynman, There's Plenty of Room at the Bottom, American Physical Society Caltech, Dec. 29, 1959

In our quest for a deeper understanding of physical and biological phenomena, we move into the "small scale" world of quantum mechanics. The rules of this world herald new types of materials and devices [1, 2, 3]. Quantum information systems and instruments of measurement promise an exponential improvement in speed and/or resolution compared to their classical counterparts. Many of these systems inherently rely on estimation for their normal operation, *e.g.*, atomic clocks, measuring electrical, thermal, and photonic characteristics, biometrics, magnetometry, and gravimetry. Some will require estimation to determine if the system is meeting performance demands and then apply a control adapted to the specific estimated error, *e.g.*, [4, 5, 6, 7, 8]. Estimation will also be used for simply gaining an understanding of observed phenomena.

Despite the promise and various laboratory successes, for many of these quantum systems *ab initio* models do not yet exist which can be used to optimize the design or determine a robust control for actual application. The only practical approach is quantum system identification – that is – identifying a model from measurements, either as an intrinsic part of their operation or in a calibration/tuning stage prior to operation. In particular, instrumentation noise, decoherence, and modeling errors are all sources of uncertainty which either separately or in combination hinder the ability of the device to meet performance demands. Finally, common to all methods of quantum system identification, as well as quantum control design [9, 10], is the computational burden imposed by the dimension of the parameter space.

2 What was proposed

We proposed to investigate a method of identification which has the potential to alleviate *all* the aforementioned problems. The question we posed was:

> *Can ℓ_1-norm minimization, which has had enormous success in signal processing for estimating a sparse variable from highly incomplete and noisy measurements, be applied to significantly improve the accuracy and efficiency of quantum system identification?*

The basic mathematical foundations for ℓ_1-norm minimization, often generally referred to as *Compressive Sensing*, can be found in [11, 12]. (A web search on Compressive Sensing will bring many tutorials and testimonials). In general, for ℓ_1-norm minimization to be effective, the underlying signal (or parameter space) must be sparse. This in turn allows for a significant reduction in the number of measurements (resources) needed for reconstructing the signal. Of course if the sparsity pattern is known then standard methods can be applied. Why it works so well is because the ℓ_1-norm is a convex heuristic for sparsity, which is not a convex function.

The analogy has been made that it is like solving a Sudoku puzzle: only a few given numbers in the grid will force a unique solution even if the grid is large.

Since compressive sensing methods can reduce resources by orders of magnitude, the benefits from a positive answer to the above question for quantum estimation would alleviate (or remove) the computational burden. Beyond this, as miraculous as it may sound, this estimation method would impact the device performance directly much as it has for digital and medical imaging, *e.g.*, [13, 14].

3 Hoped for benefits

If successful, the potential benefits include the following.

- Ancilla assisted quantum process tomography would achieve the same accuracy with a significantly smaller number of ancilla. (A quantum process tomography example in [15], repeated here in a later section, using ℓ_1-norm minimization required only 36 measurements to estimate 256 parameters compared to standard methods which require 256 measurements.)

- Quantum metrology devices which rely on entangled states to enhance accuracy would find relief in the number of entangled particles required.

- Phase estimation, which is the example posed for Phase I, is at the heart of Shor's algorithm (the quantum Fourier transform). Compressive sensing methods could significantly impact the ancilla real-estate required for the associated error-correction.

- Instrumentation limitations in both state preparation and measurement protocols would not hinder estimation efficiency.

- If Hamiltonian identification really is fast and easy, then this suggests the very important possibility of a non-qubit quantum analog computer.

4 Proposed tasks

To achieve the hoped for benefits we proposed a two-phase program. Phase was to be a theoretical study to develop the mathematical and computational tools for ℓ_1-norm minimization applied to quantum process tomography and quantum parameter estimation metrology. If Phase I was successful, then Phase II would bring in experimental components based on the mathematical and computational tools developed in Phase I. The actual scope and level of effort for Phase II will be determined in collaboration with DARPA prior to the end of the Phase I effort.

To commence we posed the following Phase I tasks:

- **Task I.1** Extend the ℓ_1-norm minimization theory to QPT. Specifically, answer the question: Is the scaling of resources linear in the number of qubits?

- **Task I.2** Develop computationally efficient ℓ_1-norm minimization algorithms which are specific for QPT.

- **Task I.3** Apply the results of Tasks I.1 and I.2 to quantum metrology. Specifically, for phase estimation in a noisy environment, answer the questions: Does the algorithm described previously converge to the correct phase within a prescribed tolerance? Does it use less resources than standard approaches? Can entangled inputs be eliminated or reduced in dimension?

As states in our proposal, it was assumed that if Phase I is successful and deemed a "GO" by DARPA, then the theory and tools developed up to that point will provide for experiments to help further develop the tools and theory. Given the emerging new concepts and software for performing quantum system identification, it was envisioned that it would be important to have a flexible working laboratory system to test the tools and refine them. Or more poetically, as Feynman put it [16]:

"*The test of all knowledge is experiment. Experiment is the sole judge of scientific 'truth.'* "

We proposed to test the capabilities of quantum system identification via ℓ_1-norm minimization with two types of experimental systems: an optical interferometer and atomic Rb, each of which provides a flexible system with well understood characteristics. Both of these systems are available at Princeton. Although we could not specify exactly the tasks for Phase II, we did propose the following task framework to be filled in after a "GO" decision has been reached.

- **Task II.1** Over the Phase II period a full battery of quantum system identification tests could be performed to benchmark the new algorithmic capabilities and provide feedback for computational improvements as well as further theoretical developments.

5 What was achieved

Early in the program, and very much earlier than anticipated, we demonstrated the effectiveness of using Compressive Sensing (CS) algorithms for Quantum process Tomography (QPT) on simulated data. (§A contains a copy of the paper.). Soon thereafter we extended CS theory for QPT to account for the restrictions imposed by quantum mechanics. We showed that for a d-dimensional system, where standard QPT requires $O(d^4)$ configurations, CS heralds $O(sd)$ configurations, where s is the sparsity level associated with the best s-sparse approximation (the actual system need not be sparse). Over the next several months – in fact almost up to the end of the originally proposed period of performance – using data obtained from a two-qubit photonic experiment at the Center for Quantum Computer Technology, Department of Physics, The University of Queensland, Australia, we demonstrated, for the first time, the use of CS for QPT, which we called CQPT for *Compressed Quantum Process Tomography*. The theoretical and experimental work was published eventually in PRL, a copy of which is contained in §B.

To summarize this: the process matrix for this 2-qubit experiment is 16x16. Taking into account the trace preserving condition, QPT requires estimating 240 real parameters. Standard methods of QPT would require at least that number of experimental configurations. Using CS methods, we obtain a 97% fidelity with 32 selected configurations and a 94% fidelity with 18 selected configurations.

These are just a few of the typical experimental results. All these conform extremely well with our early simulations, as well as being similar in character to what has been seen in audio and video processing. In these latter applications the signal sizes are significantly larger, e.g., $O(10^6)$, hence, specialized algorithms have been developed to account for the signal structure. A future effort is to develop special algorithmic structures for larger QPT.

Due to the early theory development, and especially the unanticipated and exciting early experimental success, and the time required to gather the data, some of the Phase I goals were refined, some re-defined, and some have been out of reach in the time remaining. We are very pleased about the impact of the success with "real" data which now compels some advanced and new broad and promising research directions:

- Introduction of a tailored theory and associated experiment design method for effective scaling on multi-qubit systems.
- Development of CS for Hamiltonian identification.
- Demonstrate that these ID tools can be used for control and/or device design, where in the latter case, to correct for manufacturing exigencies.

Applications for CS applied to quantum systems are just emerging in many areas. One can envision, as we have in the Phase I proposal, applications to interferometry, quantum metrology, magnetrometry, spectroscopy, and so on.

On a personal note, in my initial discussions with Dennis Healy we mused about what the potential could be for this program. He was very optimistic, but at the time, I was not ready to stick my neck out that far. Considering our success at this time, Dennis was right!

6 What remains to be done

Despite two no cost extensions, we ran out of time to thoroughly develop and test our ideas for applying compressed sensing methods to problems in Hamiltonian identification. Nonetheless we did develop a CS theory of Hamiltonian identification valid for short time scales. (This complements our previous work in Hamiltonian parameter estimation [17].) The paper on this subject will appear soon in PRA. A copy is contained in §C.

What we were ultimately after was a theory and associated computational method applicable to problems in quantum metrology or more general interferometry problems. These are essentially single parameter estimation problem. A brief summary of what we were (and are still) thinking now follows.

Figure 1 is a block-diagram operational representation of a general interferometer. Here the unknown system $\mathcal{S}(\phi_0)$ consists of a unitary $U(\phi_0)$ in channel a dependent on an unknown phase ϕ_0 followed by an unknown noise operation $\mathcal{E}$ acting on both channels a and b. The usual assumption is that the unitary is of the form $U(\phi_0) = \exp(i\phi_0 H)$ with unknown phase ϕ_0 and known Hamiltonian H [18]. Typically the range of the phase parameter is known.

The interferometric set-up is envisioned initially as a Mach-Zehnder interferometer (Fig. 2) with the addition of extra beam splitters in both arms to create photon loss as expressed schematically in Fig. 1. Figure 2 shows a schematic of the classical Mach-Zehnder interferometer for phase estimation.

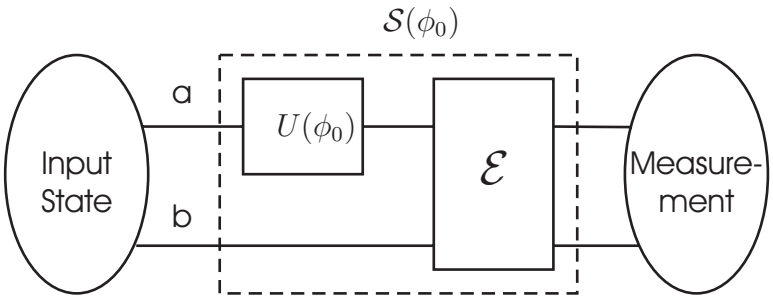
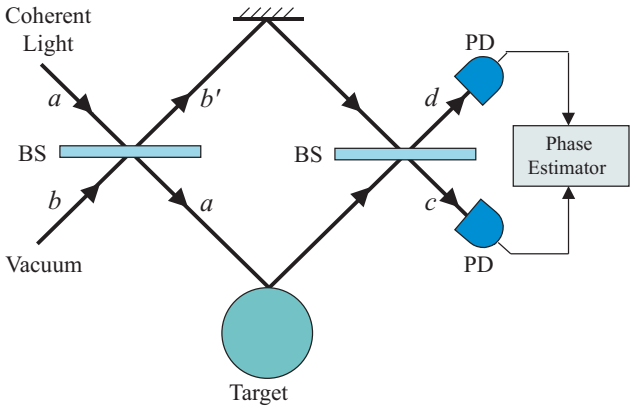


Figure 1: **General configuration for interferometric phase estimation**



A coherent light beam is split into two parts. The phase difference ϕ between the two optical arms is estimated by analyzing photon statistics of the two output beams.

BS: beam-splitter
PD: photodetector

Figure 2: **Classical Mach-Zehnder interferometer.**

For single parameter (phase) estimation the limit of theoretical accuracy in the ideal noise-free case has been examined in depth, *e.g.*, [19], [20], [21], [18], [22], [23]. These studies reveal that special preparation of the instrumentation – the probe – can achieve an asymptotic variance smaller than the Cramér-Rao lower bound, the so-called Quantum Cramér-Rao bound, or the Quantum Fisher Information (QFI). Specifically, the unique quantum property of *entanglement* can increase the parameter estimation convergence from the shot-noise limit of $1/\sqrt{N}$ to the Heisenberg limit $1/N$, which arises from the uncertainty principle [24]. In the latter case N refers to the dimension of an entangled state. The theoretical QFI, however, will not be obtained in the presence of noise, *i.e.*, decoherence. As stated in [25]:

> "Existing treatments come to the conclusion that the benefit from highly entangled states deteriorates quickly even if only a small amount of noise is present in the system ... states of this type are typically very fragile: In optical interferometry, the well-studied N00N state promises to provide Heisenberg limited sensitivity in phase estimation ... the loss of merely a single photon renders this state useless since it collapses into a product of two Fock states which can not acquire any phase information."

Fig. 6 shows a 2-qubit system where the ideal single-parameter unitary is corrupted by *ampli-*

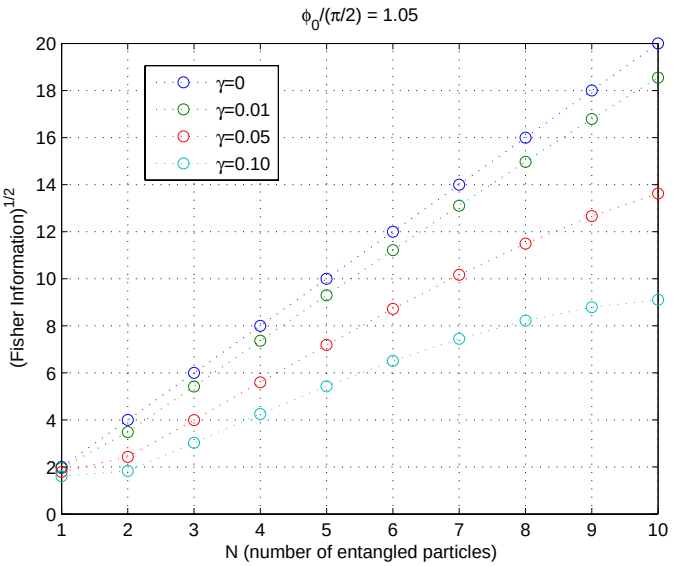


Figure 3: Fisher Information vs. number of entangled states for varying levels of amplitude damping γ.

tude damping γ. In photonic systems γ is the probability of a photon loss. In the ideal case ($\gamma = 0$) Fisher information rises linearly with the number of entangled states. However, even for a small amount of noise at the 5% level we start to see a significant loss of information. In addition to this sensitivity to noise, the QFI may also be unreachable simply because the instruments are limited, *i.e.*, not all states can be prepared and not all measurement schemes are possible, *e.g.*, [26, 17].

To alleviate these problems we proposed using a bank of estimators applied to the data, where each estimator is tuned to one of a number of finite estimates of the unknown phase parameter. For each phase estimate ϕ we will generate from the ideal unitary $U(\phi)$ an orthonormal basis set for quantum operations on the combined channel ab. A quantum process tomography will then be performed by solving an ℓ_1-norm minimization problem (compressed sensing) to obtain the phase estimate dependent process matrix. The final phase estimate is selected as the one with the smallest ℓ_1-norm of the associated process matrix.

If this approach is successful, then three significant benefits would immediately accrue. First, phase estimation would be accomplished in noisy environments. At present this is a very difficult task [25]. Secondly, the number of entangled particles might be greatly reduced. Lastly, this may also reveal an alternative to the phase estimation algorithms proposed for the Fourier transform step in many quantum algorithms [27, §5].

A Quantum process tomography via ℓ_1-norm minimization

Author Robert L. Kosut

On-line arXiv:0812.4323v2 [quant-ph]

Abstract

For an initially well designed but imperfect quantum information system, the process matrix is almost sparse in an appropriate basis. Existing theory and associated computational methods (ℓ_1-norm minimization) for reconstructing sparse signals establish conditions under which the sparse signal can be perfectly reconstructed from a very limited number of measurements (resources). Although a direct extension to quantum process tomography of the ℓ_1-norm minimization theory has not yet emerged, the numerical examples presented here, which apply .1-norm minimization to quantum process tomography, show a significant reduction in resources to achieve a desired estimation accuracy over existing methods.

4 pages

Quantum Process Tomography via ℓ_1-norm Minimization

Robert L. Kosut
SC Solutions, Sunnyvale, CA 94085 (kosut@scsolutions.com)

For an initially well designed but imperfect quantum information system, the process matrix is almost sparse in an appropriate basis. Existing theory and associated computational methods (ℓ_1-norm minimization) for reconstructing sparse signals establish conditions under which the sparse signal can be perfectly reconstructed from a very limited number of measurements (resources). Although a direct extension to quantum process tomography of the ℓ_1-norm minimization theory has not yet emerged, the numerical examples presented here, which apply ℓ_1-norm minimization to quantum process tomography, show a significant reduction in resources to achieve a desired estimation accuracy over existing methods.

Quantum process tomography (QPT) refers to the use of measured data to estimate the dynamics of a quantum system [1, 2]. Unfortunately, in the general case, the dimension of the parameter space for QPT can be prohibitive, scaling exponentially with the number of qubits. This in turn places the same burden on resources, *e.g.*, the number of applied inputs, measurement outcomes, and experiments to achieve a desired accuracy, as well as estimation computational complexity. A number of approaches have been developed to alleviate this burden. Of note are the various forms of ancilla assisted QPT (see [3] for a review), and the use of symmetrisation to estimate selected process properties [4]. Here we present a method which can be used either alone or in conjunction with any of the aforementioned approaches. The underlying premise is that for an intially well engineered design, the object that describes the quantum dynamics, the *process matrix*, will be *almost sparse* in the appropriate basis. Certainly in the ideal case of a perfect unitary channel, in the corresponding ideal basis, the process matrix is maximally sparse, *i.e.*, it has a *single* non-zero element. Since environmental interactions cannot be totally eliminated, the actual process matrix in this ideal basis will be populated with many small elements, and thus, is almost sparse.

These are the conditions under which methods using ℓ_1-norm minmization – often referred to as *Compressive Sensing* – are applicable [5, 6, 7]. Specifically, for a class of incomplete linear measurement equations ($y = Ax$, $A \in \mathbf{R}^{m\times n}$, $m \ll n$), constrained ℓ_1-norm minimization (minimize $\|x\|_{\ell_1}$ subject to $y = Ax$), a convex optimization problem, can perfectly estimate the sparse variable x. These methods also work very well for systems which do not satisfy the theoretical conditions, *i.e.*, for almost sparse variables and with measurement noise.

The underlying theory of ℓ_1 minimization shows that under certain conditions on the matrix A, to realize perfect recovery, the number of measurements, m, scales with the product of the log of the number of variables n and the sparsity. Since QPT parameters are linear in probability outcomes, and scale exponentially with the number of qubits, this approach heralds a possible linear scaling with qubits. The theory, however, has not as yet been extended to QPT. *The numerical examples here are not meant to lend support to this scaling as they are only presented for the two-qubit case.* The examples do, however, show more than an order of magnitude savings in resources over a standard constrained least-squares estimation using a complete set of measurements, *i.e.*, $\text{rank}(A) \geq n$.

The paper is organized as follows: QPT formalism is described next, followed by a discussion of the genesis of process matrix (almost) sparsity. A form of the ℓ_1 minimization for QPT is then presented followed by numerical examples and some concluding remarks.

QPT Formalism.— Recall that the state-to-state dynamics of an *open* finite-dimensional quantum system can be described in the following canonical form [1]:

$$\hat{\rho} = \textstyle\sum_{\alpha,\beta=1}^{n^2} X_{\alpha\beta}\Gamma_\alpha \rho \Gamma_\beta^\dagger \tag{1}$$

where $\rho,\ \hat{\rho} \in \mathbf{C}^{n\times n}$ are the input and output state, respectively, of dimension n, $X_{\alpha\beta}$ are the elements of the $n^2 \times n^2$ *process matrix* X, and the matrices Γ_α form an orthonormal basis set for $n \times n$ complex matrices:

$$\left\{\Gamma_\alpha \in \mathbf{C}^{n\times n} \,\middle|\, \mathbf{Tr}\ \Gamma_\alpha^\dagger \Gamma_\beta = \delta_{\alpha\beta}, \alpha, \beta = 1, \ldots, n^2 \right\} \tag{2}$$

It is assumed that the quantum system to be estimated is *completely positive and trace preserving* (CPTP). The set of feasible process matrices is then restricted to the convex set [8, 9],

$$\begin{array}{l} X \geq 0 \text{ (positive semidefinite)} \\ \sum_{\alpha,\beta=1}^{n^2} X_{\alpha\beta}\Gamma_\beta^\dagger \Gamma_\alpha = I_n \end{array} \tag{3}$$

It follows from (3) that the number of real parameters in the process matrix is $n^4 - n^2$. For q qubits $n = 2^q$, hence, scaling with parameters is exponential in the number of qubits.

Collecting data.— A common method for collecting data from a quantum system is via repeated identical experiments. Denote by $i = 1, \ldots, n_{\text{out}}$ the distinct *outcomes*, and by $k = 1, \ldots, n_{\text{cfg}}$ the experimental *configurations*, *e.g.*, any "knobs" associated with state inputs and/or measurement devices. The measurement outcomes are recorded from identical experiments in each configuration k repeated N_k times. Let N_{ik} denote the number of times out of N_k that outcome i occurred in configuration k. The QPT data are the recorded outcome counts,

$$\{N_{ik} \,|i = 1, \ldots, n_{\text{out}},\ k = 1, \ldots, n_{\text{cfg}}\} \tag{4}$$

where $N = \sum_{k=1}^{n_{\text{cfg}}} N_k = \sum_{k=1}^{n_{\text{cfg}}} \sum_{i=1}^{n_{\text{out}}} N_{ik}$ is the total number of experiments.

Estimating the process matrix.— An *empirical estimate* of the probability of measuring outcome i in configuration k can be obtained from (4) as,

$$p_{ik}^{\text{emp}} = N_{ik}/N_k \tag{5}$$

From the Born Rule the *model probability* of outcome i given configuration k with observable M_{ik} is, $p_{ik} = \mathbf{Tr}\ M_{ik}\hat{\rho}_k$, where from (1), $\hat{\rho}_k = \sum_{\alpha,\beta=1}^{n^2} X_{\alpha\beta}\Gamma_\alpha\rho_k\Gamma_\beta^\dagger$. In terms of the process matrix X, the Born rule then becomes,

$$\begin{array}{rcl} p_{ik}(X) & = & \mathbf{Tr}\, G_{ik}X \\ (G_{ik})_{\alpha\beta} & = & \mathbf{Tr}\, \Gamma_\beta^\dagger M_{ik}\Gamma_\alpha\rho_k \end{array} \tag{6}$$

The $n_{\rm out}n_{\rm cfg}$ matrices $G_{ik} \in \mathbf{C}^{n\times n}$ capture the effect of measurements in the matrix basis set (2). For each outcome i, the complete set of configurations is the combination of all these matrices and the input states: $\{\rho_k, G_{ik}\}_{k=1}^{n_{\rm cfg}}$.

A process matrix estimate can be obtained by minimizing the difference between the empirical probability estimates $p_{ik}^{\rm emp}$ and the model probabilities $p_{ik}(X)$ subject to the feasibility constraint (3). Using a "least-squares" measure of probability error leads to estimating the process matrix by solving the optimization problem:

$$\begin{array}{ll} \text{minimize} & V_{\rm LS}(X) = \sum_{i,k}\left(p_{ik}^{\rm emp} - p_{ik}(X)\right)^2 \\ \text{subject to} & X \text{ satisfies (3)} \end{array} \tag{7}$$

Because the outcomes of each experiment are independent, a maximum likelihood approach can also be considered, *i.e.*,

$$\begin{array}{ll} \text{minimize} & V_{\rm ML}(X) = -\sum_{i,k} N_{ik}\log p_{ik}(X) \\ \text{subject to} & X \text{ satisfies (3)} \end{array} \tag{8}$$

Both (7) and (8) are convex optimization problems with the optimization variables being the elements of X [8, 9]. The resulting solution (estimate) will always be CPTP (3). Unfortunately, as already mentioned, the dimension of the parameter space ($n^4 - n^2, n = 2^q$) can severely strain resources to the point of impracticality. To see this more clearly, let the linear relation in (6) between the $n_{\rm out}n_{\rm cfg}$ model probability outcomes and the n^4 elements of the process matrix be represented by an $n_{\rm out}n_{\rm cfg} \times n^4$ matrix $\mathcal{G}$, *i.e.*,

$$\vec{p} = \mathcal{G}\vec{X} \tag{9}$$

where $\vec{p}, \vec{X}$ are vectors formed from the p_{ik} and elements of X, respectively. Accounting for the n^2 linear constraints in (3), X can be recovered from either (7) or (8) to within any desired accuracy by using enough data (N in (4) sufficiently large), provided that rank$(\mathcal{G}) \geq n_{\rm out}n_{\rm cfg} \geq n^4 - n^2$. Therefore it would seem that the resources, $n_{\rm out}n_{\rm cfg}$, must also scale exponentially with the number of qubits. This, however, is not the case when the process matrix is almost sparse and where the sparsity pattern is not known[17].

Almost sparsity of the process matrix.— With no noise the ideal channel $\rho \rightarrow \hat{\rho}$ for a quantum information system is a unitary, *i.e.*, $\hat{\rho} = U\rho U^\dagger$. Let $\{\bar{\Gamma}_\alpha \in \mathbf{C}^{n\times n}\}_{\alpha=1}^{n^2}$ denote the "Natural-Basis" for matrices in $\mathbf{C}^{n\times n}$, *i.e.*, each basis matrix has a single non-zero element of one. In this basis, the process matrix associated with the ideal unitary channel has the rank-1 form, $X_{\rm ideal} = xx^\dagger$ with $x \in \mathbf{C}^{n^2}$, $x^\dagger x = n$. A singular value decomposition (SVD) gives $X_{\rm ideal} = V\,\mathrm{diag}(n, 0, \ldots, 0)V^\dagger$ with $V \in \mathbf{C}^{n\times n}$ a unitary. An equivalent process matrix can be formed from the SVD

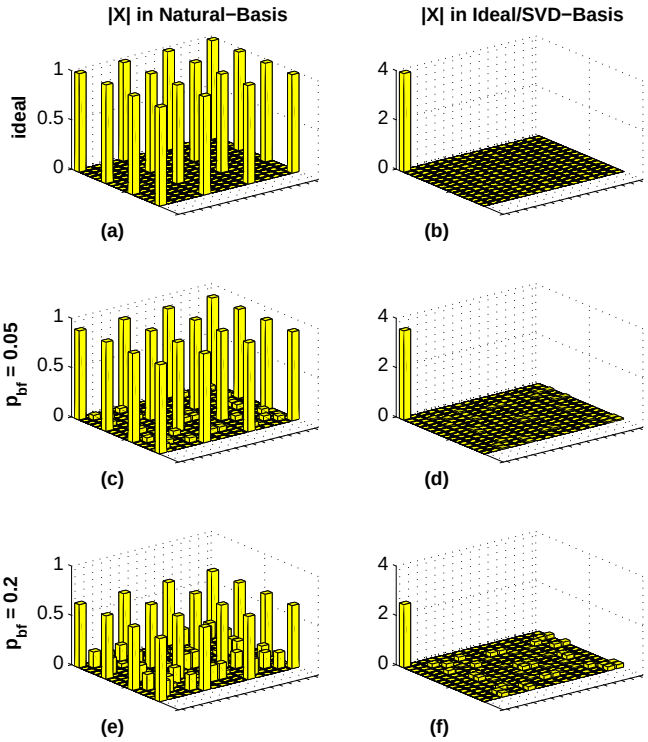


FIG. 1: Absolute values of the elements of the process matrix $X \in \mathbf{C}^{16\times 16}$ for: (a) ideal in the Natural-Basis; (b) ideal in Ideal/SVD-Basis; (c) actual ($p_{\rm bf} = 0.05$) in Natural-Basis, (d) actual ($p_{\rm bf} = 0.05$) in Ideal/SVD-Basis;(e) actual ($p_{\rm bf} = 0.2$) in Natural-Basis, (f) actual ($p_{\rm bf} = 0.2$) in Ideal/SVD-Basis.

in what is referred to here as the "Ideal/SVD-Basis," $\{\Gamma_\alpha = \sum_{\alpha'=1}^{n^2} V_{\alpha'\alpha}\bar{\Gamma}_{\alpha'} \in \mathbf{C}^{n\times n}\}_{\alpha=1}^{n^2}$, The equivalent process matrix, in this basis, denoted by $X_{\rm ideal}$, is maximally sparse with a single non-zero element, specifically, $(X_{\rm ideal})_{11} = n$. As will always be the case, the actual channel will be a perturbation of the ideal unitary. If the noise source is small then the process matrix in the nominal basis will be almost sparse.

Example: Noisy two-qubit memory.— Consider a system which is ideally a two-qubit quantum memory, thus $U = I_4, n = 4$. Suppose the actual system is a perturbation of identity by independent bit-flip errors in each channel occurring with probability $p_{\rm bf}$. For $p_{\rm bf} = 0.05$ and $p_{\rm bf} = 0.2$, the respective channel fidelities are about 0.90 and 0.64, which for quantum information processing would need to be discovered by QPT and then corrected for the device to ever work. Referring to Fig.1, in the Natural-Basis, Fig.1(a), the ideal 16×16 process matrix has 16 non-zero elements out of 256, all of magnitude one. Using the Ideal/SVD-Basis the corresponding process matrix as shown in Fig.1(b) has a *single* non-zero element of magnitude $n = 4$ – it is clearly maximally sparse. Fig.1(c)-(d) and (e)-(f), respectively, show the effect of the two $p_{\rm bf}$ levels in the two basis sets. In the Ideal/SVD-basis Fig.1(d) and (f) show that the actual (noisy) process matrices are almost sparse.

Sparsity minimization.— A known heuristic for minimizing sparsity without knowing the sparsity pattern, and also accru-

ing the benefit of using fewer resources, is to minimize the ℓ_1-norm of the vector of variables [5, 6, 9]. For QPT the equivalent ℓ_1 norm is defined here as the sum of the absolute values of the real and imaginary parts of each element of the process matrix. There are many related approaches to incorporate this norm. For example, an estimate of X can be obtained by solving the following convex optimization problem:[18]

$$\begin{array}{ll} \text{minimize} & \|X\|_{\ell_1} \equiv \sum_{\alpha,\beta=1}^{n^2} (|\mathrm{Re}\, X_{\alpha\beta}| + |\mathrm{Im}\, X_{\alpha\beta}|) \\ \text{subject to} & V(X) \leq \sigma, X \text{ satisfies (3)} \end{array} \tag{10}$$

with, *e.g.*, $V(X)$ from (7) or (8). The optimization parameter σ is used to regulate the tradeoff between fitting X to the data by minimizing $V(X)$ vs. minimizing the sparsity of X via the ℓ_1-norm. Selecting σ is often done by averaging $V(X)$ over a series of surrogates for X obtained from anticipated scenarios or iterating estimation and experiment design, *e.g.*, [8].

In the examples to follow we use the modification of (10) suggested in [7], referred to there as "ℓ_1-reweighted minimization." In this approach a weighted ℓ_1-norm is used with the weights determined iteratively. The algorithm described in [7] is:

Initialize $\sigma > 0,\ \varepsilon > 0,\ W = I_{n^4}$
Repeat

1. *Solve for* X

$$\begin{array}{ll} \text{minimize} & \|WX\|_{\ell_1} \\ \text{subject to} & V(X) \leq \sigma, X \text{ satisfies (3)} \end{array} \tag{11}$$

2. *Update weights*

$$\begin{aligned} W &= \mathrm{diag}\,(1/(|x_1| + \epsilon), \ldots, 1/(|x_{n^4}| + \epsilon)) \\ x &= \vec{X} \end{aligned} \tag{12}$$

Until *convergence* – the objective stops decreasing or a maximum number of iterations is reached.

In each of the examples to follow the procedure for QPT is: (i) solve (7) to obtain X_{ℓ_2}; (ii) set $\sigma = 1.3\, V(X_{\ell_2})$; (iii) solve the reweighting algorithm (11)-(12) for X_{ℓ_1}.

Example: QPT of noisy two-qubit memory.— For the systems from the example in Fig.1, the inputs and measurements are selected from the set of two-qubit states: $|a\rangle$, $|+\rangle = (|a\rangle + |b\rangle)/\sqrt{2}$, $|-\rangle = (|a\rangle - i|b\rangle)/\sqrt{2}$ with $a, b = 1, \ldots, 16$. Specifically, the available set of states are the 16 columns of the matrices,

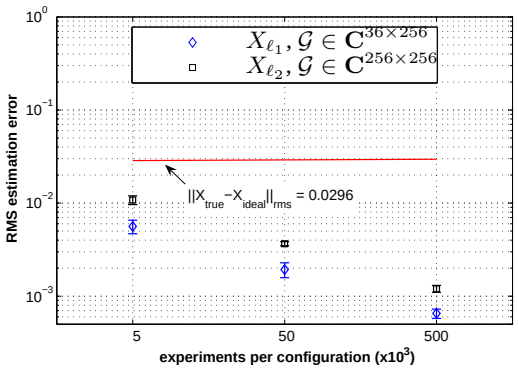
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -i & 0 & 0 & 1 & 1 & 0 \\ 0 & -i & 0 & -i & 0 & 1 \\ 0 & 0 & -i & 0 & -i & -i \end{bmatrix} \tag{13}$$

Considering only coincident input/measurement counts [10], the relevant probability outcomes (6) are,

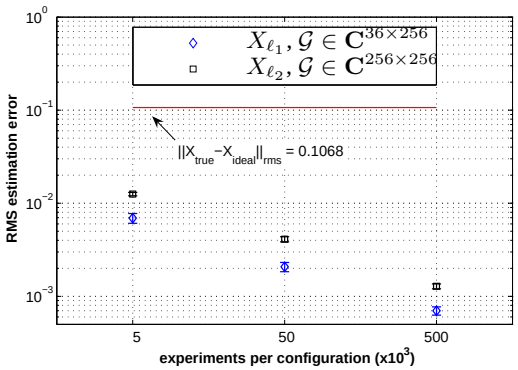
$$\begin{aligned} p_{ab}(X) &= g_{ab}^\dagger X g_{ab},\ X \in \mathbf{C}^{16\times 16} \\ (g_{ab})_\alpha &= \phi_a^\dagger \Gamma_\alpha \phi_b,\ \alpha = 1, \ldots, 16 \end{aligned} \tag{14}$$

with $\phi_a, \phi_b\ (a, b) \in \{1, \ldots, 16\}$ the selected columns of (13).

Fig.2 shows the error in estimating the process matrix $\Delta X = X_{\mathrm{true}} - X_{\mathrm{est}}$ as measured by the RMS matrix norm $\|\Delta X\|_{\mathrm{rms}} = (1/n)(\mathbf{Tr}\ \Delta X^\dagger \Delta X)^{1/2}$ vs. the number of experiments per input selected from the set (13) [19]. The results shown are from simulations described in the caption.



(a) $p_{\mathrm{bf}} = 0.05$



(b) $p_{\mathrm{bf}} = 0.20$

FIG. 2: RMS estimation error $\|X_{\mathrm{true}} - X_{\mathrm{est}}\|_{\mathrm{rms}}$ vs. number of experiments per configuration: selected columns of (13). Error bars show the deviation from 50 runs at each setting.
ℓ_2-minimization (□): $X_{\mathrm{est}} = X_{\ell_2}$ is from (7) using all 16 input/output combinations. This gives a matrix $\mathcal{G} \in \mathbf{C}^{256\times 256}$ as defined in (9) which is full rank, *i.e.*, $\mathrm{rank}(\mathcal{G}) = 256$.
ℓ_1-minimization (◊): $X_{\mathrm{est}} = X_{\ell_1}$ is from (11)-(12) using 6 inputs and 6 measurements obtained from the columns of the second matrix in (13). This gives $\mathcal{G} \in \mathbf{C}^{36\times 256}$ which is full rank, *i.e.*, $\mathrm{rank}(\mathcal{G}) = 36$.

The benefit of ℓ_1-minimization compared to the standard ℓ_2-minimization is seen most clearly with small amounts of data from highly incomplete measurements. For example, for $p_{\mathrm{bf}} = 0.05$ [Fig.2(a)], at 50×10^3 experiments per input for the 6-input/6-output configuration ($\mathcal{G} \in \mathbf{C}^{36\times 256}$) the ℓ_1 RMS estimation error is 0.0019. Compare this to the ℓ_2 error of 0.0012 at 500×10^3 experiments per input for the 16-input/16-output configuration ($\mathcal{G} \in \mathbf{C}^{256\times 256}$). The latter improvement can be attributed mostly to the 10-fold increase in the number of experiments per input. The additional resources to achieve this are significant, *i.e.*, 16 inputs for ℓ_2 vs. 6 for ℓ_1, and additionally, an increase in the *total* number of experiments from $6 \times 50 \times 10^3$ to $16 \times 500 \times 10^3$. It is certainly not intuitive that to estimate the 240 parameters of the process matrix, the clearly incomplete set of measurements using only 36 outcomes (◊ in Fig.2) could produce results not only similar

to, but for each number of experiments per input, even better than the full input case with all 256 combinations of inputs and measurements ($\Box$ in Fig.2). As seen the ℓ_1 error is about 1/2 the ℓ_2 error. Also, reweighting reduced the (unweighted) ℓ_1 error by 1/2-1/3.

Comparing the estimation errors with the error between the actual and ideal (solid lines in Fig.2) suggests that at least 50×10^3 experiments per input are needed to achieve a sufficient post-QPT error correction towards the ideal unitary. Fig.2 also reveals that the estimation errors are very similar for both levels of bit-flip error, $p_{\rm bf} \in \{0.05, 0.20\}$. This is explained by the Cramér-Rao bound which defines the asymptotic error of any unbiased estimator, *i.e.*, the RMS decays as $\Delta/\sqrt{N}$. Here Δ is effectively the error between the empirical (5) and actual (6) probabilities which by definition is of order one; this provides a reasonable fit to the data in Fig.2.

Infinite data.— With infinite data the measurements are effectively noise-free, so the empirical probability estimates are equivalent to the true probabilities. Infinite data estimates are obtained by solving (7) and (11)-(12) with the constraint $V(X) \leq \sigma$ replaced by the linear equality constraint $p_{ik}(X) = p_{ik}(X_{\rm true})$. For the numerical examples here, (14) gives the linear equality $g_{ab}^{\dagger}(X - X_{\rm true})g_{ab} = 0$.

In the examples, both X_{ℓ_1} from (11)-(12) and X_{ℓ_2} from (7) were numerically equal to $X_{\rm true}$. This is to be expected for X_{ℓ_2} because of the complete set of 256 full rank measurements. Almost sparsity makes perfect estimation possible with the highly incomplete set of 36 measurements.

The infinite data case is useful for evaluating different configuration strategies in simulation, *i.e.*, consider only those that result in a good estimate.

To stress the efficacy of ℓ_1-minimization as a heuristic for sparsity, consider replacing the ℓ_1 norm in (11)-(12) with the RMS norm $\|X\|_{\rm rms}$, which is effectively the ℓ_2 norm of $\vec{X}$. Solving the 6-input/6-output case ($\Diamond$ in Fig.2) for $p_{\rm bf} = 0.05$ with infinite data gives an RMS error of 0.11, which is considerably larger than the error between the actual and ideal of 0.03 (solid line in Fig.2(a)). The estimate gets even worse with finite data. This again emphasizes the advantage of ℓ_1 minimization for sparse signal reconstruction [5, 6].

Conclusions.— The use of the ℓ_1-norm minimization methods of Compressive Sensing [5, 6, 7] appear to apply equally well to sparse QPT. The examples of sparse process matrices presented here are meant to represent typical initial imperfect, designs. The numerical results illustrate how estimation resource tradeoffs can be obtained. Additionally, the findings suggest that QPT resources need not scale exponentially with qubits. In the ideal case, the theoretical question of showing linear scaling with sparsity using ℓ_1 minimization for QPT remains open.

Because ℓ_1 minimization uses considerably fewer resources than standard QPT, use in an on-line setting combined with optimal quantum error correction tuned to the specific QPT errors is compelling, *e.g.*, [11, 12, 13]. Another future direction is in conjunction with Hamiltonian parameter estimation. Here a bank of estimators can be applied to the data where each estimator is tuned via the Ideal/SVD-Basis to one of a number of finite samples of the unknown parameters. Such an approach may prove useful for a small number of parameters. In quantum metrology often a single uncertain parameter is to be estimated in an unknown noisy environment, *e.g.*, [14, 15].

Acknowledgments.— Thanks to A. Gilchrist, I. Walmsley, D. Lidar, H. Rabitz, and M. Mohseni for suggestions and comments. The idea of applying ℓ_1 minimization to QPT arose during discussions at [20] .

[1] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
[2] G. M. D'Ariano and P. L. Presti, Phys. Rev. Lett. **86**, 4195 (2001).
[3] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
[4] J. Emerson *et al.*, Science **317**, 1893 (2007).
[5] D. Donoho, IEEE Trans. Inform. Theory **52**, (2006).
[6] E. J. Candes, J. Romberg, and T. Tao, Comm. Pure Appl. Math. **59**, 1207 (2006).
[7] E. J. Candes, M. B. Wakin, and S. Boyd, J. Fourier Anal. and Appl. **14**, 877 (2008).
[8] R. L. Kosut, I. A. Walmsley, and H. Rabitz, quant-ph/0411093 (2004).
[9] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).
[10] J. L. O'Brien *et al.*, Phys. Rev. Lett. **93**, 080502 (2004).
[11] M. Reimpell and R. F. Werner, Phys. Rev. Lett. **94**, 080501 (2005).
[12] A. S. Fletcher, P. W. Shor, and M. Z. Win, Phys. Rev. A **75**, 012338 (2007), quant-ph/0606035.
[13] R. L. Kosut, A. Shabani, and D. A. Lidar, Phys. Rev. Lett. **100**, 020502 (2008), (arXiv:quant-ph/0703274).
[14] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
[15] U. Dorner *et al.*, quant-ph/0807.3659 (2008).
[16] M. Mohseni and A. T. Rezakhani, quant-ph/arXiv:0805.3188 (2008).
[17] A known sparsity pattern can arise from the underlying dynamics, thereby inherently increasing QPT efficiency [16].
[18] There are many alternatives to (10), *e.g.*, minimize $\|X\|_{\ell_1} + \lambda V(X)$ subject to X satisfies (3) or minimize $V(X)$ subject to $\|X\|_{\ell_1} \leq s$, X satisfies (3) .
[19] The number of experiments per input/measurement configuration here is chosen uniformly. An optimal (non-uniform) choice which minimizes the Cramér-Rao lower bound can be cast as a convex optimization problem [8].
[20] *Workshop on Quantum Estimation: Theory and Practice*, Aug. 25-30, 2008, Perimeter Institute, Waterloo, Canada.

B Efficient measurement of quantum dynamics via compressive sensing

Authors A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, A. G. White

Journal Phys. Rev. Lett. 106, 100401 (2011)

On-line arXiv:0910.5498v2 [quant-ph]

Abstract

The resources required to characterize the dynamics of engineered quantum systems – such as quantum computers and quantum sensors – grow exponentially with system size. Here we adapt techniques from compressive sensing to exponentially reduce the experimental configurations required for quantum process tomography. Our method is applicable to processes that are nearly sparse in a certain basis and can be implemented using only single-body preparations and measurements. We perform efficient, high-fidelity estimation of process matrices of a photonic two-qubit logic gate. The database is obtained under various decoherence strengths. Our technique is both accurate and noise robust, thus removing a key roadblock to the development and scaling of quantum technologies.

9 pages

Efficient Measurement of Quantum Dynamics via Compressive Sensing

A. Shabani,[1] R. L. Kosut,[2] M. Mohseni,[3] H. Rabitz,[1] M. A. Broome,[4] M. P. Almeida,[4] A. Fedrizzi,[4] and A. G. White[4]

[1]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*
[2]*SC Solutions, Sunnyvale, California 94085, USA*
[3]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[4]*Center for Engineered Quantum Systems and Center for Quantum Computation and Communication Technology, School of Mathematics and Physics, The University of Queensland, QLD 4072, Australia*

(Received 5 November 2009; revised manuscript received 14 November 2010; published 7 March 2011)

The resources required to characterize the dynamics of engineered quantum systems—such as quantum computers and quantum sensors—grow exponentially with system size. Here we adapt techniques from compressive sensing to exponentially reduce the experimental configurations required for quantum process tomography. Our method is applicable to processes that are nearly sparse in a certain basis and can be implemented using only single-body preparations and measurements. We perform efficient, high-fidelity estimation of process matrices of a photonic two-qubit logic gate. The database is obtained under various decoherence strengths. Our technique is both accurate and noise robust, thus removing a key roadblock to the development and scaling of quantum technologies.

DOI: 10.1103/PhysRevLett.106.100401 PACS numbers: 03.65.Wj, 03.65.Yz, 03.67.Lx

Understanding and controlling the world at the nanoscale—be it in biological, chemical or physical phenomena—requires quantum mechanics. It is therefore essential to characterize and monitor realistic complex quantum systems that inevitably interact with typically uncontrollable environments. One of the most general descriptions of the dynamics of an open quantum system is a quantum map—typically represented by a process matrix [1]. Methods to identify this matrix are collectively known as quantum process tomography (QPT) [1,2]. For a d-dimensional quantum system, they require $O(d^4)$ experimental configurations: combinations of input states, on which the process acts, and a set of output observables. For a system of n qubits—two level quantum systems—$d = 2^n$. The required physical resources hence scale exponentially with system size. Recently, a number of alternative methods have been developed for efficient and selective estimation of quantum processes [3]. However, full characterization of quantum dynamics of comparably small systems, such as an 8-qubit ion trap [4], would still require over a billion experimental configurations, clearly impractical. So far, process tomography has therefore been limited by experimental and off-line computational resources, to systems of 2 and 3 qubits [5–7].

Here we adapt techniques from compressive sensing to develop an experimentally efficient method for QPT. It requires only $O(s \log d)$ configurations if the process matrix is s compressible in some known basis, i.e., it is nearly sparse in that it can be well approximated by an s-sparse process matrix. This is usually the case, because engineered quantum systems aim to implement a unitary process which is maximally sparse in its eigenbasis. In practice, as observed in liquid-state NMR [8], photonics [5,9,10], ion traps [11], and superconducting circuits [6], a near-unitary process will still be nearly sparse in this basis, and still compressible. The near sparsity is due to few dominant system environment interactions. This is more apparent for weakly decohering systems [12].

We experimentally demonstrate our algorithm by estimating the 240 real parameters of the process matrix of a canonical photonic two-qubit gate, Fig. 1, from a reduced number of configurations. From just 18 and 32 configurations, we obtain fidelities of 94% and 97% with process matrices obtained from an overcomplete set of all 576 available configurations.

Compressive sensing provides methods for compression of information carried by a large-size signal into a significantly smaller one along with efficient convex optimization algorithms to decipher this information [13]. Originally developed to exploit compressible features of

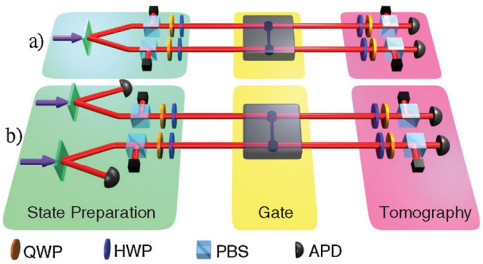


FIG. 1 (color online). Experimental scheme. Two-photon inputs were prepared with either (a) a high-rate, nonscalable, two-photon source or (b) a low-rate, scalable, four-photon source. The qubits are encoded using polarization, as described in the text. The quantum process is a photonic entangling gate. A measurement configuration is defined as some combination of state preparation and an observable, implemented here with quarter- and half-wave plates (QWP, HWP), polarizers (PBS), and photon detectors (APD). For details see [19].

audio and video signals, compressive sensing is now applied to: simulations of compressive sensing for QPT [14], ghost imaging [15], and state tomography for low-rank density matrices [16]. The latter provides a quadratic reduction of physical resources from d^2 for standard tomography to $O(rd\log^2 d)$ for a density matrix of rank r with the added advantage that rank is basis independent. Recently, this method has been useful in efficient state tomography of one-dimensional systems approximated by matrix product states [17].

Under reasonable assumptions, a quantum map on a d-dimensional space has the general representation [1],

$$\mathcal{S}(\rho) = \sum_{\alpha,\beta=1}^{d^2} \chi_{\alpha\beta}\Gamma_\alpha \rho \Gamma_\beta^\dagger \tag{1}$$

where χ, the $d^2 \times d^2$ process matrix, is positive semidefinite, $\chi \geq 0$, and trace preserving, $\sum_{\alpha,\beta}\chi_{\alpha\beta}\Gamma_\beta^\dagger\Gamma_\alpha = I_d$, with $\{\Gamma_\alpha\}$ an orthonormal matrix basis set, $\mathrm{Tr}(\Gamma_\beta^\dagger\Gamma_\alpha) = \delta_{\alpha\beta}$. Note that sparsity is a property of the map representation not the map itself. Data is collected by preparing an ensemble of identical systems in one of the states $\{\rho_1, \ldots, \rho_k\}$, inputting them to the process χ, and then measuring an observable M from the set $\{M_1, \ldots, M_\ell\}$. For a pair (ρ, M), the outcome will be $y_{M,\rho} = \mathrm{Tr}(\mathcal{S}(\rho)M)$. If the experiment is repeated for all configurations, i.e., (ρ_i, M_i), $i = 1, \ldots, m = k\ell$, the relation between the vector of outcomes $y = [y_{M_1,\rho_1}, \ldots, y_{M_m,\rho_m}]^T$ and the true process matrix, denoted by χ_0, can be represented by a linear map $y = \Phi\vec{\chi}_0$, where $\vec{\chi}_0$ is the vectorized form of the process matrix χ_0 and Φ is an $m \times d^4$ matrix of coefficients of the form $\mathrm{Tr}(\Gamma_\alpha\rho_i\Gamma_\beta^\dagger M_i)/\sqrt{m}$.

In general, estimating a sparse process matrix with an unknown sparsity pattern from an underdetermined set of linear equations ($m < d^4$) would seem highly unlikely. Compressive sensing, however, tells us that this can be done by solving for χ from the convex optimization problem:

$$\text{minimize } \|\vec{\chi}\|_{\ell_1} \text{ subject to } \|y - \Phi\vec{\chi}\|_{\ell_2} \leq \varepsilon, \tag{2}$$

and positive-semidefinite and trace-preserving conditions as defined above. The parameter ε quantifies the level of uncertainty in the measurements, that is, we observe $y = \Phi\chi_0 + w$ with $\|w\|_{\ell_2} \leq \varepsilon$. From [18], recovery via (2) is ensured if (i) the matrix Φ satisfies the restricted isometry property:

$$1 - \delta_s \leq \frac{\|\Phi\vec{\chi}_1(s) - \Phi\vec{\chi}_2(s)\|_{\ell_2}^2}{\|\vec{\chi}_1(s) - \vec{\chi}_2(s)\|_{\ell_2}^2} \leq 1 + \delta_s \tag{3}$$

for all s-sparse $\chi_1(s)$, $\chi_2(s)$ process matrices; (ii) the isometry constant $\delta_{2s} < \sqrt{2} - 1$ and (iii) the number of configurations $m \geq C_0 s\log(d^4/s)$. Under these conditions, the solution $\chi^\star$ of (2) satisfies,

$$\|\vec{\chi}^\star - \vec{\chi}_0\|_{\ell_2} \leq \frac{C_1}{\sqrt{s}}\|\vec{\chi}_0(s) - \vec{\chi}_0\|_{\ell_1} + C_2\varepsilon \tag{4}$$

where $\chi_0(s)$ is the best s-sparse approximation of χ_0 and C_0, C_1, C_2 are constants on the order of $O(\delta_s)$, see [19]. The restricted isometry property states that two s-sparse process matrices $\chi_1(s)$ and $\chi_2(s)$ can be distinguished if their relative distance is nearly preserved after the measurements. If the measurements are noise free, $\varepsilon = 0$, and χ_0 is s sparse, $\chi_0 = \chi_0(s)$, then the right-hand side of (4) is zero leading to perfect recovery, $\chi^\star = \chi_0$. Otherwise the solution tends to the best s-sparse approximation of the process matrix plus the additional term due to measurement error ε. If for an n-qubit QPT with $d = 2^n$ the conditions of the above analysis are satisfied, then the number of experimental configurations m scales linearly with sn, specifically, $m \geq C_0 s(4n\log 2 - \log s) = O(sn)$. In [19], using the measure concentration properties of random matrices, following the arguments in [20], we show that if Φ is constructed from random input states $\{\rho_i\}$, and random observables $\{M_i\}$, then the restricted isometry in (3) holds with high probability. Also a test is presented to certify the sparsity assumption.

A nearly sparse process matrix can be recovered from an exponentially smaller number of measurement outcomes to within the bounds of (4) by solving (2). We now test our algorithm experimentally against standard QPT on a two-qubit gate under a range of decoherence—and thus sparsity—conditions. We used a photonic controlled-phase, CZ, gate, Fig. 1 where the qubits are encoded in orthogonal polarization states of single photons ($|H\rangle$, horizontal, and $|V\rangle$, vertical). We performed full process tomography [5,9,10] of the gate with both two-photon and four-photon arrangements, preparing 16 pairwise combinations of the 4 input states $\{|H\rangle, |V\rangle, |D\rangle, |R\rangle\}$ and, for each input, measuring 36 two-qubit combinations of the observables $\{|H\rangle, |V\rangle, |D\rangle, |A\rangle, |R\rangle, |L\rangle\}$, where $|D(A)\rangle = (|H\rangle \pm |V\rangle)/\sqrt{2}$ and $|R(L)\rangle = (|H\rangle \pm i|V\rangle)/\sqrt{2}$. These 576 input-output configurations represent an overcomplete set which allows the best possible estimate of the quantum process, denoted χ_{576} [5].

The compressed quantum process tomography (CQPT) estimate of the 16×16 process matrix, χ_m, is obtained by solving (2) with $y = C_{\mathrm{sel}}p$ and $\Phi = C_{\mathrm{sel}}G$ where p is the vector of 576 experimental probabilities corresponding to each of the 576 configurations, G is the 576×256 matrix obtained from all the configurations with the basis set $\{\Gamma_\alpha\}$, and C_{sel} is the $m \times 576$ matrix corresponding to taking a selection of $m \leq 576$ of all possible configurations. The basis set is obtained from the singular-value decomposition of the ideal CZ gate: the process matrix in this basis is maximally sparse with a single nonzero (1, 1) element. The measurement error bound ε in (2) is chosen to be just slightly larger than $\sqrt{m}\sigma$, where σ is the minimum feasible root-mean-square level obtained from (2) using all configurations, i.e., with $C_{\mathrm{sel}} = I_{576}$. We quantify decoherence using the process purity, $\mathcal{P} = \mathrm{Tr}(\chi_m^2/d^2)$, which varies from 0 for a completely decohering channel, to 1 for a unitary process: in our experiment we used six

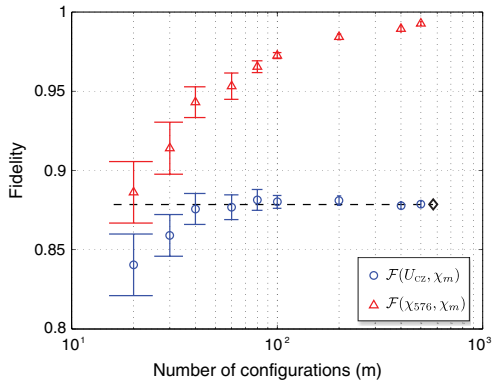


FIG. 2 (color online). Process fidelities vs number of input-output configurations, for each compressive QPT estimate, χ_m, in the gate basis of the ideal CZ gate for the lowest measured noise level, $\mathcal{P} = 0.91$. The dashed line shows the fidelity of the full estimate $\mathcal{F}(U_{cz}, \chi_{576}) = 0.89$ (black diamond). Error bars are obtained by solving (2) for 50 different random combinations of m inputs and observables.

decoherence levels (see [19] for details), giving purities of $\{0.62, 0.74, 0.77, 0.79, 0.86, 0.91\} \pm 0.01$.

Figure 2 shows, for the lowest decoherence level, the process fidelities [5] versus the number of randomly selected configurations, m. Each process matrix, $\{\chi_m\}$, is obtained by solving (2). We use the fidelity between (i) the compressive measurement and the ideal, $\mathcal{F}(U_{cz}, \chi_m)$; and (ii) the compressed and optimal measurements, $\mathcal{F}(\chi_{576}, \chi_m)$. Note that as m increases the fidelity with the ideal converges to the value of 0.89 obtained from χ_{576}; likewise, the fidelity with the full estimate converges to unity. Similar plots exist for every level of decoherence, with fidelities reduced accordingly.

We have so far used random selections of probabilities from the full data set, which allows us a comprehensive test of CQPT. Experiments, however, do not yield probabilities but physical quantities, e.g., count rates. To date, algorithms for more efficient state [16] or process tomography have assumed probabilities as a starting point. Since normalization is an issue to some extent in all physical architectures, it will be necessary to investigate the robustness and scalability of algorithms for real-world experiments.

For our photonic two-qubit gate, which is lossy and intrinsically probabilistic, the probabilities were obtained by normalizing counts using a full basis set of observables extracted from all measurements, I_{576}. Having sufficient configurations to allow for normalization necessarily imposes limits on CQPT efficiency: for low m, we are restricted in how random our selections can be. (Details and some permissible configurations in [19]). As an example, Fig. 3 shows process matrices reconstructed via CQPT from just one of these configurations compared to the respective full data estimates. We used 32 combinations of the 16 inputs $\{|H\rangle, |V\rangle, |D\rangle, |R\rangle\}^{\otimes 2}$ and 2 observables $\{|R\rangle|I\rangle, |I\rangle|R\rangle\}$, where I is the identity. The agreement is excellent as one can see from the fidelities and the correct reproduction of imaginary elements—which are ideally zero. Another striking feature is that we obtain highly faithful reconstructions of a nonlocal process using only local measurements [2].

A further crucial test is whether CQPT enables us to locate errors and implement necessary corrections: a common example is identifying local rotations that move the process closer to the ideal. By optimizing $\mathcal{F}(U_{cz}, \chi_{32})$, we calculated local corrections to χ_{32}; applying them to the full estimate χ_{576}, $\mathcal{F}(U_{cz}, \chi_{576})$ improved, on average, over all decoherence levels, by 4.1%. This is very close to the average 4.9% improvement obtained by calculating and applying local corrections *directly* to χ_{576}. Even a low-configuration CQPT estimate of a noisy process therefore enables improvements.

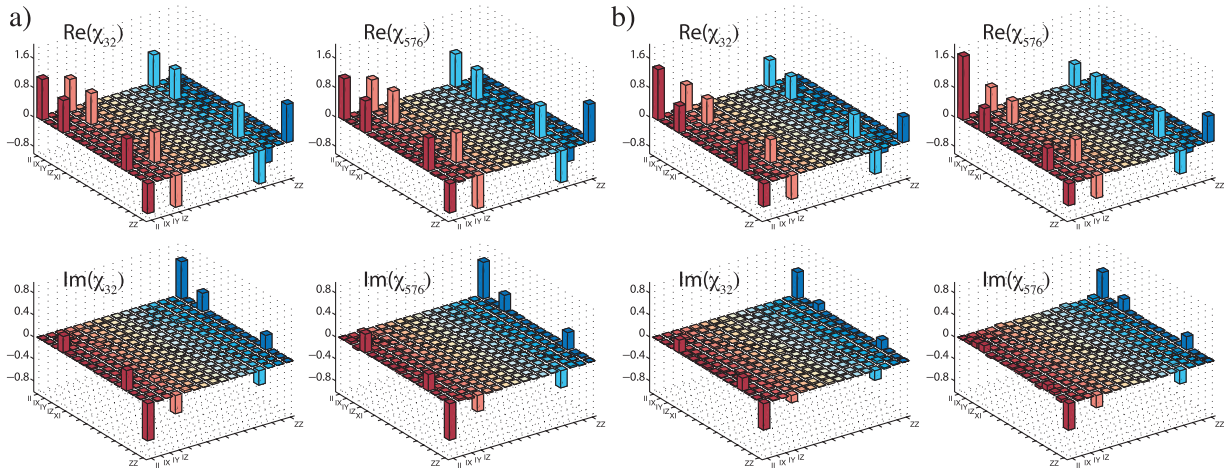


FIG. 3 (color online). Real and imaginary process matrix elements in the Pauli basis for the CQPT estimate χ_{32}, 32 configurations (left) vs full data estimate χ_{576}, 576 configurations (right) for (a) a low noise, two-photon experiment, $\mathcal{P} = 0.91$, and (b) a high-noise, four-photon experiment, $\mathcal{P} = 0.62$. The CQPT reconstructions have fidelities, $\mathcal{F}(\chi_{576}, \chi_{32})$, of 95% and 85%, respectively. The CQPT estimation accuracy is excellent for low noise, and reliable even for high noise, see [19] for more details.

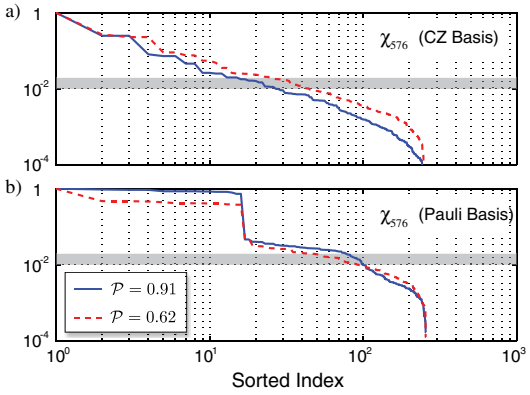


FIG. 4 (color online). Absolute values of the 256 process matrix elements of χ_{576} for our lowest and highest noise level, sorted by relative magnitude [with respect to the (1, 1) element] in the CZ basis (top) and the Pauli basis (bottom). The error threshold, which indicates the required number of configurations, is shown in grey.

That high-fidelity estimates are obtained by CQPT can be understood from the error bound (4) which shows that the CQPT estimate tends towards the best s-sparse approximation of the true process, χ_{576}. Figure 4 shows the process matrix elements, sorted by relative magnitude, for low- and high-noise levels, in two basis sets. The s-sparse approximation levels indicated in (4) are reached where the matrix elements drop below the error threshold (0.01–0.02). For the corresponding m, we can therefore expect a successful, high-fidelity, CQPT reconstruction. In the CZ basis, the plots show that for low noise, $s \in [20, 30]$, which correlates well with the fidelities in Fig. 2; for high noise $s \in [40, 60]$. Although the process matrix is still somewhat sparse in the Pauli basis (Fig. 3). Figure 4(b) indicates that ~100 configurations are needed to obtain an estimate of comparable quality. Furthermore, the sorted magnitude values in the CZ basis decay exponentially, which is sufficient to declare the process matrix s compressible, see, e.g., [21,22]. Intriguingly, this exponential decay is a signature of model-based compressive sensing where the scaling goes from $m = O(s \log(d/s))$ to $m = O(s)$ [22]. This demands further investigation, since it appears that QPT fits this framework, particularly when the process matrix is expanded in the ideal basis corresponding to the unitary design goal.

Our experimental results are supported numerically by simulations of a 2-qubit process as well as simulation studies for 3- and 4-qubit systems which show the same type of compressibility, see [19]. Applying CQPT to larger systems will require careful attention to classical postprocessing which—as in QPT—scales exponentially. The standard software we used here (see [19]), can easily handle 2- and 3-qubit CQPT systems. For larger systems, more specialized software can increase speed by orders of magnitude, see, e.g., [21].

A number of research directions arise from this work: incorporating knowledge of model structure properties; tightening the bounds on scaling laws; understanding how near-sparsity s and rank r vary with system dimension, d; pursuing highly efficient convex-computational algorithms; and selection of optimal configurations. Compressive tomography techniques can also be applied to quantum metrology and Hamiltonian parameter estimation: for example, estimating selective properties of biological or chemical interest in molecular systems and nanostructures with typically sparse Hamiltonians [23].

We thank J. Romberg and S. Jafarpour for discussions. We acknowledge funding by the ARC Discovery and Federation Fellowship programs and an IARPA-funded US Army Research Office contract. R. L. K. and H. R are supported by DARPA Grant No. FA9550-09-1-0710 administered through AFOSR.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
[2] J. B. Altepeter *et al.*, Phys. Rev. Lett. **90**, 193601 (2003); M. Mohseni and D. Lidar, Phys. Rev. Lett. **97**, 170501 (2006); M. Lobino *et al.*, Science **322**, 563 (2008); M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
[3] J. Emerson *et al.*, Science **317**, 1893 (2007); A. Bendersky *et al.*, Phys. Rev. Lett. **100**, 190403 (2008); M. Branderhorst *et al.*, New J. Phys. **11**, 115010 (2009).
[4] H. Haffner *et al.*, Nature (London) **438**, 643 (2005).
[5] J. O'Brien *et al.*, Phys. Rev. Lett. **93**, 080502 (2004).
[6] R. Bialczak *et al.*, Nature Phys. **6**, 409 (2010).
[7] T. Monz *et al.*, Phys. Rev. Lett. **102**, 040501 (2009).
[8] A. M. Childs, I. L. Chuang, and D. W. Leung, Phys. Rev. A **64**, 012314 (2001); N. Boulant *et al.*, Phys. Rev. A **67**, 042322 (2003); Y. Weinstein *et al.*, J. Chem. Phys. **121**, 6117 (2004).
[9] M. W. Mitchell *et al.*, Phys. Rev. Lett. **91**, 120402 (2003).
[10] Y. Nambu and K. Nakamura, Phys. Rev. Lett. **94**, 010404 (2005).
[11] M. Riebe *et al.*, Phys. Rev. Lett. **97**, 220407 (2006).
[12] M. Mohseni and A. T. Rezakhani, Phys. Rev. A **80**, 010101 (2009); A. G. Kofman and A. N. Korotkov, Phys. Rev. A **80**, 042103 (2009).
[13] E. Candes and M. Wakin, IEEE Signal Process. Mag. **25**, 21 (2008).
[14] R. L. Kosut, arXiv:0812.4323v1.
[15] O. Katz *et al.*, Appl. Phys. Lett. **95**, 131110 (2009).
[16] D. Gross *et al.*, Phys. Rev. Lett. **105**, 150401 (2010).
[17] M. Cramer *et al.*, Nature Commun. **1**, 149 (2010).
[18] E. J. Candes, C.R. Seances Acad. Sci. Ser. 1 **346**, 589 (2008).
[19] See supplemental material at http://link.aps.org/supplemental/10.1103/PhysRevLett.106.100401.
[20] R. Baraniuk *et al.*, Constr. Aprox. **28**, 253 (2008).
[21] D. Needell and J. A. Tropp, Appl. Comput. Harmon. Anal. **26** 301 (2009).
[22] R. G. Baraniuk *et al.*, IEEE Trans. Inf. Theory **56**, 1982 (2010).
[23] J. Yuen-Zhou *et al.*, arXiv:1006.4866.

Online material: Efficient measurement of quantum dynamics via compressive sensing

A. Shabani,[1] R. L. Kosut,[2], M. Mohseni[3], H. Rabitz[1], M. A. Broome[4], M. P. Almeida[4], A. Fedrizzi[4], and A. G. White[4]

[1]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA,*
[2]*SC Solutions, Sunnyvale, California 94085, USA,* [3]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA,*
[4]*Centre for Engineered Quantum Systems and Centre for Quantum Computer and Communication Technology, School of Mathematics and Physics, The University of Queensland, QLD 4072, Australia*

APPENDIX A: NORMS

Definitions of the norms used throughout the paper. For a vector $x \in \mathbf{C}^n$,

$$\begin{array}{rcl} \|x\|_{\ell_2} & = & \sqrt{x^\dagger x} = \sqrt{\sum_{i=1}^n |x_i|^2} \\ \|x\|_{\ell_1} & = & \sum_{i=1}^n |x_i|. \end{array} \tag{A1}$$

For a matrix $A \in \mathbf{C}^{m\times n}$ with $\mathbf{rank}(A) = r \leq \min\{m,n\}$ and singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$,

$$\begin{array}{lrcl} \text{Induced } \ell_2 \text{ norm} & \|A\|_2 & = & \sup_{\|x\|_{\ell_2}=1} \|Ax\|_{\ell_2} = \sigma_1 \\ \text{Frobenius norm} & \|A\|_{\text{fro}} & = & \sqrt{\mathbf{Tr}(A^\dagger A)} = \sqrt{\sum_{i=1}^r \sigma_i^2} \\ \text{Nuclear norm} & \|A\|_\star & = & \mathbf{Tr}(\sqrt{A^\dagger A}) = \sum_{i=1}^r \sigma_i \end{array} \tag{A2}$$

In the main theorem of CQPT we evaluate the distance between the vectorized form of two process matrices, $\|\vec{\chi}_2 - \vec{\chi}_1\|_{\ell_1}$. It is interesting to see what is the relation between this distance and a more natural measure of distance between two maps. A commonly used definition of distance between two quantum maps $\mathcal{S}_2$ and $\mathcal{S}_1$ is $||\mathcal{S}_2 - \mathcal{S}_1|| = d\sup_{A,B} |tr[A\mathcal{S}_2(B) - A\mathcal{S}_1(B)]|$ for all matrices A and B such that $\|A\|_{\text{fro}} = \|B\|_{\text{fro}} = 1$. This can be equivalently expressed as $d\sup_{A,B} |\text{tr}[(\tau_2 - \tau_1)A \otimes B]|$ where τ_α ($\alpha = 1, 2$) is the Jamiolkowski state equivalence of $\mathcal{S}_\alpha$ defined as $\frac{1}{d}\sum_{i,j} |i\rangle\langle j| \otimes \mathcal{S}_\alpha(|i\rangle\langle j|)$. The distance $||\mathcal{S}_2 - \mathcal{S}_1||$ is upper bounded by $d||\tau_2 - \tau_1||_*$ [4]. Using the orthonormality properties of the basis Γ_α one can show $||\tau_2 - \tau_1||_{\text{fro}} = ||\chi_2 - \chi_1||_{\text{fro}}$. Finally we can use $||A||_* \leq \sqrt{r}||A||_{\text{fro}}$ (r is the rank of matrix A) to find $\|\vec{\chi}_2 - \vec{\chi}_1\|_{\ell_1} \geq \frac{1}{\sqrt{r}d}||\mathcal{S}_2 - \mathcal{S}_1||$.

APPENDIX B: RESTRICTED ISOMETRY PROPERTY FROM A CONCENTRATION INEQUALITY

A common approach to establish the restricted isometry property (RIP), Eqn.(3) in the paper, for a matrix $A \in \mathbf{C}^{m\times n}$ with $m < n$ is by introducing randomness in the elements of this matrix. This approach benefits from measure concentration properties of random matrices. For QPT for the measurment matrix $\Phi \in \mathbf{C}^{m\times d^4}$ in Eqn.(2) of the paper, we show how to achieve this with random preparation of the intial states and a random selection of the measurement operators. The proof is based on the results in [5] which show that if Φ is a random matrix which satisfies the concentration property,

$$\mathbf{Pr}\left\{ |\, \|\Phi x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2 \,| \geq \delta_s \|x\|_{\ell_2}^2 \right\} \leq 2e^{-2mC_3(\delta_s)}, \tag{B1}$$

for all $x \in \mathbf{C}^{d^4}$, where $\delta_s \in (0,1)$ and $C_3(\delta_s)$ only depends on δ_s, then Φ satisfies the RIP,

$$(1-\delta_s)\|x_s\|_{\ell_2}^2 \leq \|\Phi x_s\|_{\ell_2}^2 \leq (1+\delta_s)\|x_s\|_{\ell_2}^2 \tag{B2}$$

for all s-sparse $x_s \in \mathbf{C}^{d^4}$. This version of RIP is equivalent to Eqn.(3) in the paper.

In classical signal processing, each element of the Φ matrix can be independently selected from a random distribution such as Gaussian or Bernoulli. For QPT there is no freedom for random independent selection of every element of the Φ matrix. However, as described in the paper, the *rows* of Φ can be independently and randomly selected. To see this, recall that for each experimental configuration we can initialize the system randomly in a state $\rho \in \{\rho_i \in \mathbf{C}^{d\times d}\}_{i=1}^k$ and then measure an observable M randomly chosen from $\{M_j \in \mathbf{C}^{d\times d}\}_{j=1}^\ell$. The corresponding matrix Φ then has $m = k\ell$ independent random rows $\{\phi_i^\dagger \in \mathbf{C}^{1\times N}\}_{i=1}^m$ with correlated elements of each row since they are functions of the same M and ρ. Observe, however, that although Φ is a random matrix, because it is constructed from quantum states and observables of a finite dimensional system, it is bounded. As a consequence, $\forall x \in \mathbf{C}^{d^4}$, we get,

$$\begin{array}{rcl} (w_\ell/m)\|x\|_{\ell_2}^2 \;\leq\; x^\dagger(\phi_i\phi_i^\dagger)x \;\leq\; (w_u/m)\|x\|_{\ell_2}^2 \\ \ell\|x\|_{\ell_2}^2 \;\leq\; \mathbf{E}\|Ax\|_{\ell_2}^2 \;\leq\; u\|x\|_{\ell_2}^2 \end{array} \tag{B3}$$

where $\mathbf{E}$ denotes expectation with respect to Φ and w_u, w_ℓ, u, ℓ are constants. Next we apply,

Hoeffding's concentration inequality *Let $v_1, ..., v_m$ be independent bounded random variables such that v_i falls in the interval $[a_i, b_i]$ with probability one. Then for $S = \sum_i v_i$ and any $t > 0$ we have,*

$$\begin{array}{rcl} \mathbf{Pr}\{S - \mathbf{E}(S) \geq t\} \;\leq\; e^{-2t^2/\sum_i (b_i-a_i)^2} \\ \mathbf{Pr}\{S - \mathbf{E}(S) \leq -t\} \;\leq\; e^{-2t^2/\sum_i (b_i-a_i)^2} \end{array} \tag{B4}$$

In our problem $v_i = |\phi_i^\dagger x|^2$ and $S = \|\Phi x\|_{\ell_2}^2$. From the above inequalities and the relations in (B3) we find $\forall t_+, t_- > 0$ and $\forall x$,

$$\begin{array}{rcl} \mathbf{Pr}\left\{S - u\|x\|_{\ell_2}^2 \geq t_+\right\} &\leq& \mathbf{Pr}\{S - \mathbf{E}(S) \geq t_+\} \\ &\leq& e^{-2t_+^2/(w_u - w_\ell)^2} \\ \mathbf{Pr}\left\{S - l\|x\|_{\ell_2}^2 \leq -t_-\right\} &\leq& \mathbf{Pr}\{S - \mathbf{E}(S) \leq -t_-\} \\ &\leq& e^{-2t_-^2/(w_u - w_\ell)^2} \end{array} \tag{B5}$$

The choice of $t_+ = (\delta_s + 1 - u)\,\|x\|^2_{\ell_2}$ and $t_- = (\ell - 1 + \delta_s)\,\|x\|^2_{\ell_2}$ in the above inequalities yields

$$\mathbf{Pr}\left\{|\,\|\Phi x\|^2_{\ell_2} - \|x\|^2_{l_2^m}\,| \geq \delta_s\,\|x\|^2_{\ell_2}\right\} \leq 2e^{-2m(\delta_s+\epsilon)^2/(w_u-w_\ell)^2} \tag{B6}$$

with $\epsilon = \min\{1-u, \ell-1\}$. We also need t_+ and t_- to be positive that imposes the condition $1-\delta_s < \ell \leq u < 1+\delta_s$. Since the obervable M can be scaled by any real factor, a sufficient condition is $u/\ell < (1+\delta_s)/(1-\delta_s)$.

Next we reproduce the connection between the measure concentration (B6) and restricted isometry as demonstrated in [5]: Let X_s be a set of vectors with cardinality s: $\#(X_s) = s$. We choose a set $Y \subset X_s$ such that $\|y\|_{\ell_2} = 1$ for all $y \in Y$, we have $\min_{y\in Y}\|x-y\|_{\ell_2} \leq \delta_s/4$ for all $x \in X_s$. The cardinality of such a set Y can always be chosen to be smaller than $(12/\delta_s)^s$ [6]. There from (B6) we find

$$\mathbf{Pr}\left\{|\,\|\Phi y\|^2_{\ell_2} - 1| \geq \delta_s/2\right\} \leq 2(12/\delta_s)^s e^{\frac{-2m(\delta_s/2+\epsilon)^2}{(w_u-w_\ell)^2}}$$

or equivalently $1-\delta_s/2 \leq \|\Phi y\|^2_{\ell_2} \leq 1+\delta_s/2$ holds with probability exceeding

$$P = 1 - 2(12/\delta_s)^s \exp(-2m(\delta_s/2+\epsilon)^2/(w_u-w_\ell)^2).$$

Define z to be the smallest number such that $\|\Phi x'\|_{\ell_2} \leq 1+z$ for all x' with $\|x'\|_{\ell_2} = 1$. For a vector $y \in Y$ we have,

$$\|\Phi x'\|_{\ell_2} \leq \|\Phi y\|_{\ell_2} + \|\Phi(x'-y)\|_{\ell_2} \leq 1 + \frac{\delta_s}{2} + (1+z)\frac{\delta_s}{4}$$

from which it follows that $z < \delta_s$, for any $0 < \delta_s < 1$. In a similar fashion we can prove $1-\delta_s \leq \|\Phi x'\|_{\ell_2}$. This completes the proof that RIP (B2) holds with probability exceeding P for all $x \in X_s$. The number of sets X_s with $\#X_s = s$ is $\binom{N}{s} \leq (eN/s)^s$. Therefore RIP fails to be satisfied with probability $2\exp(-2m(\delta_s/2+\epsilon)^2/(w_u-w_\ell)^2 + s[log(eN/s)+log(12/\delta_s)])$. For a sufficiently small constant C_0, if $C_0 s \leq m/log(N/s)$, we can find a constant $0 < C_3$ such that the probability of a failure of RIP becomes smaller than $\exp(-C_3 m)$ provided that $C_3 \leq 2m(\delta_s/2+\epsilon)^2/(w_u - w_\ell)^2 - s[log(eN/s)+log(12/\delta_s)]$. This guaranteed exponentially small chance of RIP failure is the key to the logarithmic scaling of the resources in CQPT. If RIP is satisfied the l_1 norm minimization algorithm works to find a sparse solution. Here we proved that by increasing the number of configurations m would exponentially decrease the chance of RIP failure. This completes the connection between the concentration measure (B6) and the restricted isometry property.

APPENDIX C: PERFORMANCE OF THE ALGORITHM

In Ref. [2], the accuracy of the ℓ_1-norm minimization problem is given by (C2). The parameters C_1 and C_2 are explicitly given in terms of the isometry constant δ_s:

$$C_1 = \frac{2 + (2\sqrt{2}-2)\delta_s}{1-(\sqrt{2}+1)\delta_s}, \quad C_2 = \frac{4\sqrt{1+\delta_s}}{1-(\sqrt{2}+1)\delta_s} \tag{C1}$$

To present all the distances based on l_1-norm we can use $||y||_{l_1} \leq ||y||_{l_2} \leq \sqrt{D}||y||_{l_1}$, for a D-dimensional vector y and obtain the algorithm performance as

$$\|\vec{\chi}^\star - \vec{\chi}_0\|_{\ell_1} \leq \frac{C_1 d^2}{\sqrt{s}}\,\|\vec{\chi}_0(s) - \vec{\chi}_0\|_{\ell_1} + d^2 C_2 \varepsilon \tag{C2}$$

However the performance inequality presented in the paper has a tighter bound.

APPENDIX D: SPARSITY ASSUMPTION CERTIFICATION

A test to certify the sparsity assumption can be concluded from (C2) and the probability of RIP being satisfied exceeding $1 - e^{-mC_3(\delta_s)}$ for m configurations. Suppose an estimate χ_m is obtained for m configurations. If the measure $\|\chi_{m+1} - \chi_m\|_{\ell_1}$, which quantifies an incremental improvement in the estimated process matrix, converges toward zero for a polynomially large m, the sparsity assumption is certified.

APPENDIX E: NORMALIZATION AND PRECISION ISSUES

In the formulation of CQPT a random selection of the expectation values y_{M_i,ρ_i} are not available in our experiment. Due to photon loss the detector counts are not conclusive, hence, a complete set of counts corresponding to a complete set of observables is required to produce meaningful expectation values y_{M_i,ρ_i}. A solution to this problem is to limit the measurements to few-body observables. For k-body measurements a total number of 2^k complementary observables need to be measured. Since m, the number of measurements, is exponentially small we can choose k limited to few-body operators, $k = k_{max}$, and even single-body as we did in the experiment. For a fully random selection of observables, the total number of measurements m will be increased by a constant factor 2^k_{max}. Still this number is exponentially small. This redundancy, however, can be avoided by using the outcomes of all 2^k observables. This selection scheme is not fully random, rather it is a deterministic-random way of choosing observables.

As discussed in the paper, random selections of probabilities from the full data set, although exhibiting results which are entirely consistent with compressive sensing theory, are inconsistent with how data is actually collected in this kind of standard photonic experiment. In practice we are limited to measure few-body observables. For low m, the configurations must allow for normalisation, i.e. we are restricted in how random our low-number selections can be. A selection of some of these permissible configurations are shown in Table I. Here we see some of the remarkable results promised by the theory of compressed sensing, *e.g.*, a 98% fidelity from 32 configurations and a 94% fidelity from only 18 configurations.

Another issue to consider is experimental precision. The expectation values of k-body observables of random states reduce for a larger k. This implies the need for a larger number

Inputs	Observables	m	$\mathcal{F}(\chi_{576},\chi_m)$	$\mathcal{F}(U_{CZ},\chi_m)$
HVDR$^{\times 2}$	HVDARL$^{\times 2}$	576	1	0.88
HVDR$^{\times 2}$	{RI,IR}	32	0.98	0.89
	{DI,ID}		0.97	0.87
HVDR$^{\times 2}$	RL$^{\times 2}$	64	0.95	0.86
	DAxDA		0.95	0.86
VDR$^{\times 2}$	{RI,IR}	18	0.94	0.86
	{DI,ID}		0.93	0.88
VDR$^{\times 2}$	RL$^{\times 2}$	36	0.94	0.87
	DA$^{\times 2}$		0.94	0.84

TABLE I: Fidelity assessment of some selected configurations that are available in our experiment.

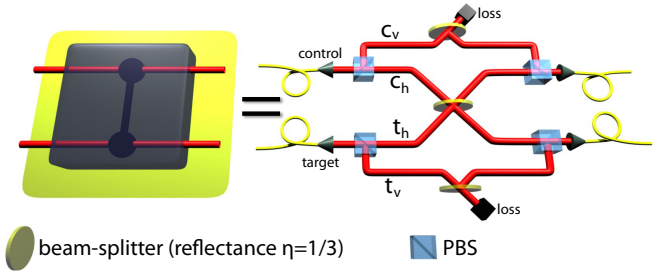


FIG. 1: Detailed representation of the CZ-gate in dual rail encoding. Each qubit is represented by two paths, one for each logical basis state, $|0\rangle = |H\rangle$ and $|1\rangle = |V\rangle$ [9].

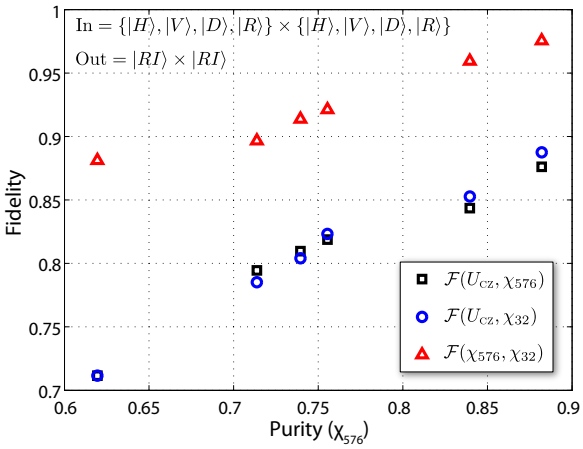


FIG. 2: Fidelities vs. purities for $m = 32$ corresponding to the configurations in Table I.

of statistical samples. Fortunately, this issue is not a problem for our scheme since we can take k as small as we want, as discussed above.

APPENDIX F: CLASSICAL POSTPROCESSING

The estimation results computed from the experimental data were all obtained by solving equation 2 in the main text by using "off-the-shelf" MATLAB based software. Specifically, we used *YALMIP* to call the convex solver *SDPT3* [7, 8]. On a standard desktop it takes about 2 sec of CPU-time to solve (2) for the full 576 configuration set. This software can handle 3-qubit systems but it is more advisable to migrate to more specialized software where orders of magnitude speed increases are possible, *e.g.*, [3].

APPENDIX G: EXPERIMENTAL DETAILS

The quantum gate used in the experiment is a photonic controlled-phase gate, Fig. 1 [9]. It is based on a single partially polarising beam splitter (PPBS), having different reflectivity, $\eta_V = \frac{1}{3}$, $\eta_H = 0$, for the horizontal and the vertical polarisation of input photons. Due to two-photon interference, the input state $|VV\rangle$ undergoes a π phase shift $|VV\rangle \rightarrow -|VV\rangle$ whenever the two photons leave the PPBS through different output ports. Correct operation of the gate is signalled by a coincidence detection in these output modes; the gate is thus probabilistic, with a success probability of $1/9$.

The gate acts on photonic qubits created via spontaneous parametric downconversion (SPDC). Downconversion is intrinsically a random process: consequently the created states contain small amounts of higher-order emission—e.g. $|22\rangle$ as well as the desired $|11\rangle$—which appear as decoherence in a quantum process [10, 11]. The ratio of higher order terms to the desired photon pair number increases with the pair creation probability, which in turn is proportional to the pump laser power. Once can therefore—to some extent—control the decoherence in a process via the laser power.

In order to cover a comprehensive range of decoherence, we performed six experiments with 2-photon states directly created via a single SPDC emission, and one experiment with 4-photon states created in two independent SPDC sources, where one photon of each SPDC process was used as a trigger. The latter experiment is more representative of large-scale systems, where independent photon sources will be required. It has significantly reduced count rates, and reduced two-photon interference between photons in the quantum gate due to both the pump-induced decoherence and group-velocity mismatch [11], reflected in the low purity of the process in this case of 0.62.

Typical count rates for 2-photon experiments are 2000 coincident counts per second, full QPT, building up reasonable statistics, takes about 2.5 hours; in contrast, 4-photon experiments have much lower rates, 1 four-fold coincidence per second, and take 2 *days*. The 32-configuration CQPT reduces tomography times to 8 minutes and 2.6 hours respectively: a clear advantage.

Fig. 2 shows the effect of varying laser pump power on CQPT estimation accuracy for one of the single-observable configurations from Table I. Specifically for the 32 configurations arising from all combinations of the 16 inputs HVDRx-HVDR and 2 outcomes {RI,IR}. As pump power increases, the process purity, as measured by $\mathbf{Tr}(\chi^2_{576})/16$ decreases;

effectively the signal to noise ratio deteriorates. As might be expected, the worst-case fidelity decreases with process purity. The estimated channel fidelity is however remarkably robust, staying very close to the actual channel fidelity.

APPENDIX H: SIMULATION RESULTS

QPT is performed by solving Eqn. (2) in the paper with noise-free experiments ($\epsilon = 0$) for a system designed to be a 2-qubit quantum Fourier transform (QFT) with unitary representation $U_{\rm qft} \in \mathbb{C}^{4\times4}$, which interacts with an unknown environment via the total *constant* (time-independent) Hamiltonian, $H{=}I_e \otimes H_{\rm qft}{+}\gamma\tilde{H}$ with $\tilde{H}$ randomly selected and normalized to $\|\tilde{H}\|{=}1$; γ is thus the interaction magnitude. The simulated system $\chi_{\rm sim} \in \mathbb{C}^{16\times16}$ is extracted via the partial trace over the environment for $\gamma \in \{0.5, 1.0, 1.25\}$. Each of these induces a fidelity with respect to the ideal unitary, $\mathcal{F}(U_{\rm qft}, \chi_{\rm sim}) \in \{0.70, 0.80, 0.95\}$ The estimates from Eqn. (2) in the paper are obtained in the singular value decomposition (SVD)-basis [Γ_α in Eqn.(1) of the paper of the ideal QFT. The process matrix of the ideal unitary in this basis is maximally sparse with the single non-zero 1,1-element equal to $n = 4$ [1]. The environmental interactions make the process matrix nearly sparse, *i.e.*, compressible.

To form the measurement matrix $\Phi \in \mathbb{C}^{m\times256}$, we randomly generated 4 and 16 input pairs $|\psi_1\rangle \otimes |\psi_2\rangle$ and 2, 4, and 6 random selections from the *single-body* Pauli observables $\{IX, IY, IZ, XI, YI, ZI\}$. This gives 6 configurations with $m \in \{8, 16, 24, 32, 64, 96\}$, for which $u/\ell \approx 1.3$ ensuring $\delta \approx 0.13$. Fig. 3 shows the fidelities $\mathcal{F}(\chi_m, \chi_{\rm sim})$ of the reconstructed estimates χ_m and the simulated process matrices $\chi_{\rm sim}$ for all 18 combinations of m and interaction magnitudes γ.

These results arise from the relative sparsity of the process matrix in the SVD-basis of the ideal QFT. Fig. 4 shows 3D bar plots of the real and imaginary elements of the true and estimated process matrices for $m = 64$, $\mathcal{F}(U_{\rm qft}, \chi_{\rm sim}) = 0.70$, and $\mathcal{F}(\chi_{64}, \chi_{\rm sim}) = 0.93$. In the SVD-basis (row 2), the true process matrix exhibits the expected large 1,1-element with the remaining elements much smaller by comparison. The estimated channel fidelity is 0.71.

In Fig. 3, $\mathcal{F}(\chi_m, \chi_{\rm sim})$ (white bars) trends to increase with m, more so for $\mathcal{F} = 0.7$ than for $\mathcal{F} = 0.95$, and rises a bit sharply at different m values. Just as for the experimental results, this can be connected to the actual sparsity of the simulated process matrices. Figure 5, just like Fig.4 in the main text, shows the absolute sorted process matrix elements relative to the 1,1-element. Where each plot crosses the threshold of 0.02, we see that the number of elements above this value increases with decreasing decoherence γ. If these are taken as the s-sparse approximation levels indicated in the theory, Eqn. (4) in the paper, then (approximately) $s \in \{30, 50, 100\}$ correspond to $\mathcal{F}(U_{\rm qft}, \chi_{\rm sim}) \in \{0.95, 0.80.0.70\}$. This correlates well with how $\mathcal{F}(\chi_m, \chi_{\rm sim})$ varies with resources m.

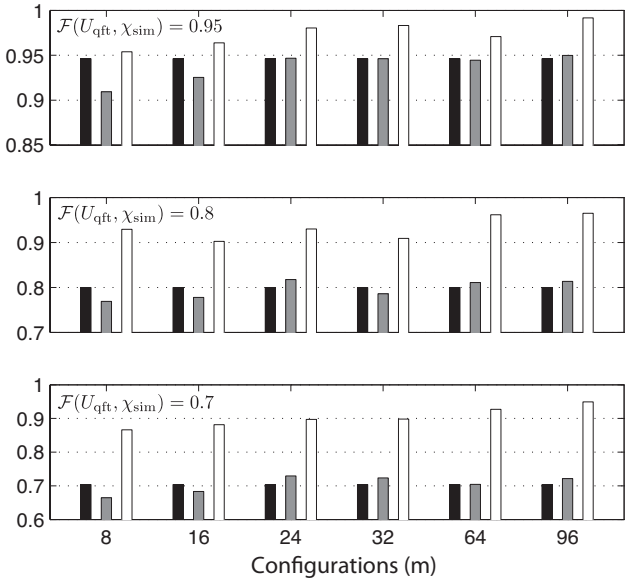


FIG. 3: Fidelities vs. configurations for each process matrix estimate χ_m from Eqn. (2) in the paper in the SVD basis of the ideal QFT unitary. Black bars: simulated compared to ideal process $\mathcal{F}(U_{\rm qft}, \chi_{\rm sim})$. Gray bars: estimate compared to ideal $\mathcal{F}(U_{\rm qft}, \chi_m)$. White bars: estimate compared to simulated process $\mathcal{F}(\chi_m, \chi_{\rm sim})$.

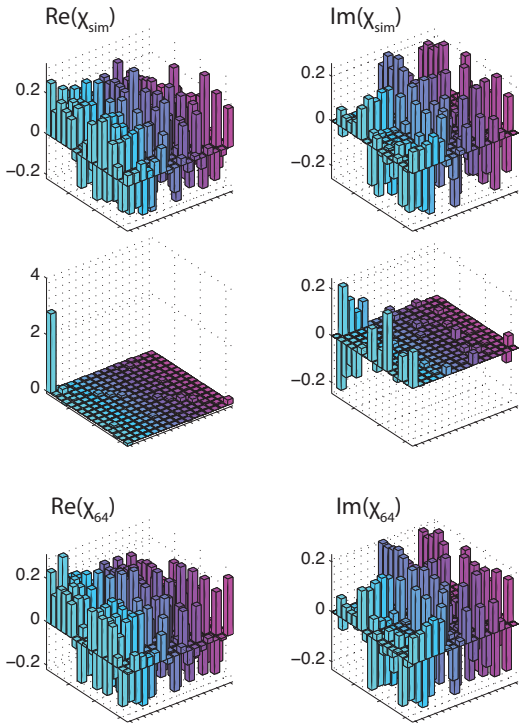


FIG. 4: Real and imaginary χ elements for $m{=}64$, $\mathcal{F}(U\text{qft}, \chi_{\rm sim}) = 0.71$, $\gamma{=}1.25$. Row 1: True process matrix in the natural basis. Row 2: True process matrix in SVD-basis of ideal unitary. Row 3: Estimated process matrix projected to the natural basis , $\mathcal{F}(U_{\rm qft}, \chi_m) = 0.71$.

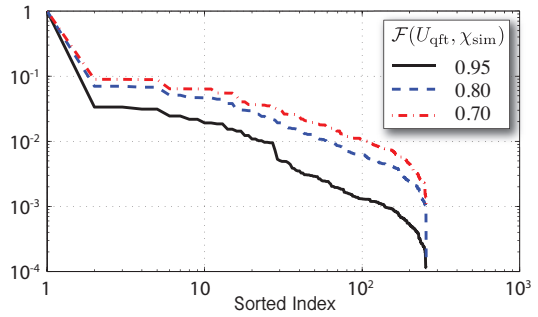


FIG. 5: Absolute values of the 256 process matrix elements of $\chi^{\rm true}_{\rm svd} \in \mathbf{C}^{16\times 16}$ sorted by relative magnitude (with respect to the 11-element) for $\mathcal{F}(U_{\rm qft}, \mathcal{S}^{\rm true}) \in \{0.95, 0.80, 0.70\}$.

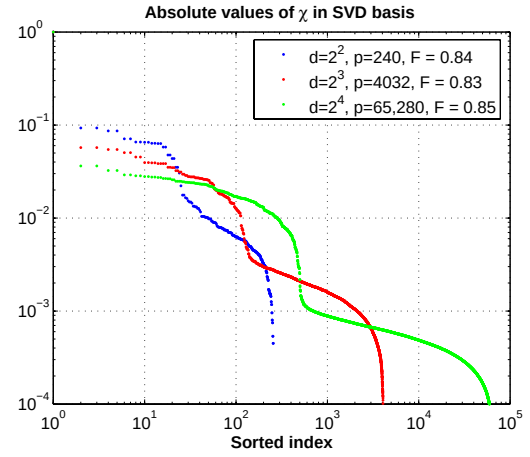


FIG. 6: Absolute values of the process matrix elements sorted by relative magnitude (with respect to the 11-element) all in the ideal SVD basis (in this case for an identity operator on the system) for three cases: **blue** $\chi_{\rm sim} \in \mathbf{C}^{16\times 16}$ with $\mathcal{F}(I_4, \chi_{\rm sim}) = 0.84$; **red** $\chi_{\rm sim} \in \mathbf{C}^{64\times 64}$ with $\mathcal{F}(I_8, \chi_{\rm sim}) = 0.83$; **green** $\chi_{\rm sim} \in \mathbf{C}^{256\times 256}$ with $\mathcal{F}(I_{16}, \chi_{\rm sim}) = 0.85$;

APPENDIX I: BEYOND 2-QUBITS

Standard QPT scales exponentially, thus for 3 and 4 qubits the number of required experimental configurations is, respectively 4,032 and 65,280. As we have shown theoretically, experimentaly, and lastly via the previous simulations, CQPT shows quite a different scaling. Fig. 6 shows the absolute values, sorted by relative magnitude, of the process matrices arising from a random selection of a perturbed system near identity, *i.e.*, a quantum memory, corresponding to similar fidelities. The process matrices elements are shown in a basis corresponding to the ideal identity. Again taking 0.01 as a threshold we see that for 2-qubits we get $m \approx 30$ which is similar to our experimental results and those supported by the plots in Figures 4 in the main text and here in 5. Fig. 6 predicts for 3-qubits $m \approx 100$, and for 4-qubits $m \approx 300$. These simulation results show first that the process matrices are compressible, and in addition are consistent with the experimental results in Fig. 4 in the main text. To actually perform the estimaton, that is solve Eqn. (2) in the paper, as previously mentioned, requires specialized compressed sensing algorithms optimized for speed and efficiency, *e.g.*, [3].

[1] R. L. Kosut, arXiv:0812.4323v1[quant-ph] (2008).
[2] E. J. Candes, Compte Rendus de l'Academie des Sciences, Paris, Serie I **346**, 589 (2008).
[3] D. Needell and J. A. Tropp, Appl. Comp. Harmonic Anal. pp. 301–321 (2008).
[4] M. M. Wolf,and J. I. Cirac, Commun. Math. Phys. **147**, 279 (2008).
[5] R. Baraniuk *et al.*, Constructive Approximation **28**, 253 (2008).
[6] G. G. Lorentz *et al.*, Grundlehren der Mathematischen Wissenschaften **304** (1996).
[7] J. Lofberg, in *Proceedings of the CACSD Conference* (Taipei, Taiwan, 2004).
[8] K. C. Toh *et al.*, (2004), http://www.math.nus.edu.sg/∼mattohkc/sdpt3.html.
[9] N. K. Langford *et al.*, Phys. Rev. Lett. **95**, 21 (2005);R. Okamoto *et al.* , Phys. Rev. Lett. **95**, 210506 (2005); N. Kiesel *et al.*, Phys. Rev. Lett. **95**, 210505 (2005).
[10] M. Barbieri *et al.*, Journal of Modern Optics **56**, 209 (2009).
[11] T. Weinhold *et al.*, Preprint at arXiv:0808.0794 (2008).

C Efficient estimation of nearly sparse many-body quantum Hamiltonians

Authors A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, H. Rabitz

Journal Physical Review A, to appear.

On-line arXiv:1002.1330v2 [quant-ph]

Abstract

We develop an efficient and robust approach for quantum measurement of nearly-sparse many-body quantum Hamiltonians based on the method of compressive sensing. This work demonstrates that with only O(s log(d)) experimental configurations, consisting of random local preparations and measurements, one can estimate the Hamiltonian of a d-dimensional system, provided that the Hamiltonian is nearly s-sparse in a known basis. The classical post-processing is a convex optimization problem on the total Hilbert space which is generally not scalable. We numerically simulate the performance of this algorithm for three- and four-body interactions in spin-coupled quantum dots and atoms in optical lattices. Furthermore, we apply the algorithm to characterize Hamiltonian fine structure and unknown system-bath interactions.

8 pages

Efficient estimation of nearly sparse many-body quantum Hamiltonians

A. Shabani,[1] M. Mohseni,[2] S. Lloyd,[2,3] R. L. Kosut,[4] and H. Rabitz[1]

[1]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544*
[2]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139*
[3]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139*
[4]*SC Solutions, Sunnyvale, CA 94085*

We develop an efficient and robust approach to Hamiltonian identification for multipartite quantum systems based on the method of compressed sensing. This work demonstrates that with only $\mathcal{O}(s\log(d))$ experimental configurations, consisting of random local preparations and measurements, one can estimate the Hamiltonian of a d-dimensional system, provided that the Hamiltonian is nearly s-sparse in a known basis. We numerically simulate the performance of this algorithm for three- and four-body interactions in spin-coupled quantum dots and atoms in optical lattices. Furthermore, we apply the algorithm to characterize Hamiltonian fine structure and unknown system-bath interactions.

PACS numbers:

I. INTRODUCTION

The dynamical behavior of multipartite quantum systems is governed by the interactions amongst the constituent particles. Although, the physical or engineering considerations may specify some generic properties about the nature of quantum dynamics, the specific form and the strength of multi-particle interactions are typically unknown. Additionally, quantum systems usually have an unspecified interaction with their surrounding environment. In principle, one can characterize quantum dynamical systems via "quantum process tomography" (QPT) [1–8]. However, the relationship between relevant physical properties of a system to the information gathered via QPT is typically unknown. Alternatively, knowledge about the nature of inter- and intra- many-body interactions within the system and/or its environment can be constructed by identifying a set of (physical or effective) Hamiltonian parameters generating the dynamics [9–18]. Currently, a scalable approach for efficient estimation of a full set of Hamiltonian parameters does not exist.

The dynamics of a quantum system can be estimated by observing the evolution of some suitable test states. This can be achieved by a complete set of experimental configurations consisting of appropriate input states and observables measured at given time intervals. Knowledge about the dynamics may then be reconstructed via inversion of the laboratory data by fitting a set of dynamical variables to the desired accuracy. Estimating Hamiltonian parameters from such a procedure faces three major problems: (1) The number of required physical resources grows exponentially with the degrees of freedom of the system [1–8]. (2) There are inevitable statistical errors associated with the inversion of experimental data [1–8]. (3) The inversion generally involves solving a set of nonlinear and non-convex equations, since the propagator is a nonlinear function of Hamiltonian parameters [9–18]. The first two problems are always present with any form of quantum tomography, but the last problem is specific to the task of Hamiltonian identification as we wish to reconstruct the generators of the dynamics. Many quantum systems involve two-body local interactions, so the goal is often to estimate sparse Hamiltonians with effectively a polynomial number of unknown parameters. Unfortunately, quantum state and process tomography cannot readily exploit this potentially useful feature.

The highly nonlinear feature in the required inversion of laboratory data was studied in Ref.[9] in which closed-loop learning control strategies were used for the Hamiltonian identification. In that approach one estimates the unknown Hamiltonian parameters by tailoring shaped laser pulses to enhance the quality of the inversion. Identification of time-independent (or piece-wise constant) Hamiltonians have been studied for single-qubit and two-qubit cases [13, 14] to verify the performance of quantum gates. Estimation of these Hamiltonians is typically achieved via monitoring the expectation values of some observable, e.g. concurrence, which are time periodic functions. Through Fourier transform of this signal the identification task is reduced to finding the relative location of the peaks and heights of the Fourier spectrum [13, 14]. Bayesian analysis is another method proposed for robust estimation of a two-qubit Hamiltonian [15]. The difficulty with these methods is then scalability with the size of the system. A symmetrization method for efficient estimation of the magnitude of effective two-body error generators in a quantum computer was studied in [16] by monitoring quantum gate average fidelity decay. Recently, it was demonstrated that direct or selective QPT schemes could be used for efficient identification of short-time behavior of sparse Hamiltonians [17] assuming controllable two-body quantum correlations with auxiliary systems and the exact knowledge of the sparsity pattern. Another scheme for the determination of the coupling parameters in a chain of interacting spins with restricted controllability was introduced in Ref. [18].

In this work, inspired by recent advances in classical signal processing known as *compressed sensing* [19], we use random *local* input states and measurement observables for efficient Hamiltonian identification. We show how the difficulties with the nonlinearity of the equations can be avoided by either a short time or a perturbative treatment of the dynamics. We demonstrate that randomization of the measurement observables enables compressing the extracted Hamiltonian information into a exponentially smaller set of outcomes. This is accomplished by a generalization of compressed sensing to

utilize random matrices with correlated elements. This approach is applicable for Hamiltonians that are nearly sparse in a known basis with an arbitrary unknown sparsity pattern of parameters. The laboratory data can then be inverted by solving a convex optimization problem. This algorithm is highly tolerant to noise and experimental imperfections. The power of this procedure is illustrated by simulating three- and four-body Hamiltonians for neutral atoms in an optical lattice and spin-coupled quantum dot systems, respectively. Furthermore, we directly apply the algorithm to estimate Hamiltonian fine structure and characterize unknown system-bath interactions for open quantum systems.

II. QUANTUM DYNAMICAL EQUATIONS

The time evolution of a quantum system in a pure state is governed by the Shrödinger equation, $d\left|\psi(t)\right\rangle/dt = -iH\left|\psi(t)\right\rangle$. The solution of this equation for a time-independent Hamiltonian can be simply expressed as $\left|\psi(t)\right\rangle = \exp(-itH)\left|\psi(0)\right\rangle$. In principle, the Hamiltonian of the system H can be estimated by preparing an appropriate set of test states $\{\left|\psi_k\right\rangle\}$ and measuring the expectation value of a set of observables $\{M_j\}$ after the system has evolved for a certain period of time. The expectation value of these observables can be expressed as

$$p_{jk} = \langle M_j\rangle_{\psi_k} = \left\langle\psi_k\right| e^{itH} M_j e^{-itH}\left|\psi_k\right\rangle \qquad (1)$$

Equation (1) implies that the experimental outcomes $\{p_{jk}\}$ are nonlinear functions of the Hamiltonian parameters. To avoid the difficulties of solving a set of coupled nonlinear equations we consider the short time behavior of the system. Monitoring the short time dynamics of the system is valid when the relevant time scales of the system evolution satisfy $t \ll K^{-1}$ where, for positive operator-valued measure (POVM) operators $\{M_j\}$, the constant K equals $2||H||_{spec}$. The general expression of K is given in appendix B, also see appendix A for definition of the norms. This yields the linearized form of the Eq. (1)

$$p_{jk} = \left\langle\psi_k\right| M_j\left|\psi_k\right\rangle + it\left\langle\psi_k\right|[H, M_j]\left|\psi_k\right\rangle + \mathcal{O}(K^2t^2) \quad (2)$$

The linear approximation contains enough information to fully identify the Hamiltonian and the higher order terms do not provide additional information. The short-time approximation implies prior knowledge about the system dynamical time-scale or the order of magnitude of $||H||_{spec}$. This prior knowledge can be available from generic physical and engineering considerations. For example, in solid-state quantum devices the time-scale of single qubit rotations is typically on the order of 1-10 ns. The switching time for exchange interactions varies among different solid-state systems from 1ps to 100ps, (for more details see appendix B.)

We expand the Hamiltonian in an orthonormal basis $\{\Gamma_\alpha\}$, where $\mathrm{Tr}(\Gamma_\alpha^\dagger\Gamma_\beta) = d\delta_{\alpha,\beta}$: $H = \sum_\alpha h_\alpha\Gamma_\alpha$. Here d is the dimension of the Hilbert space. In this representation the Hamiltonian parameters are the coefficients h_α. The expanded form of the above affine equation (2) is

$$\bar{p}_{jk} = it\sum_\alpha \left\langle\psi_k\right|[\Gamma_\alpha, M_j]\left|\psi_k\right\rangle h_\alpha \qquad (3)$$

Here we introduce the experimental outcomes as $\bar{p}_{jk} = p_{jk} - \left\langle\psi_k\right| M_j\left|\psi_k\right\rangle$, since $\left\langle\psi_k\right| M_j\left|\psi_k\right\rangle$ is *a priori* known. The relation (3) corresponds to a single experimental configuration (M_j,$|\psi_k\rangle$). For a d-dimensional system, the total number of Hamiltonian parameters h_α is d^2. Thus, one requires the same number of experimental outcomes, p_{jk} that leads to d^2 linearly independent equations. For a system of n qubits, this number grows exponentially with n as $d = 2^{2n}$. In order to devise an efficient measurement strategy we will focus on physically motivated *nearly sparse* Hamiltonians.

A Hamiltonian H is considered to be s-sparse if it only contains s non-zero parameters $\{h_\alpha\}$. More generally, a Hamiltonian H is termed nearly s-sparse, for a threshold η, if at most s coefficients h_α ($H = \sum h_\alpha\Gamma_\alpha$) have magnitude greater than ηh_{max} where $h_{max} = \max(h_\alpha)$. By definition, the sparsity is basis dependent. However, for local interactions, the basis in which the Hamiltonian is sparse is typically known from physical or engineering considerations.

III. COMPRESSED HAMILTONIAN ESTIMATION

Our algorithm is based on general methods of so-called compressed sensing that recently have been developed in signal processing theory [19]. Compressed sensing allows for condensing signals and images into a significantly smaller amount of data, and recovery of the signal becomes possible from far fewer measurements than required by traditional methods.

Compressed sensing has two main steps: encoding and decoding. The information contained in the signal is mapped into a set of laboratory data with an exponentially smaller representation. This compression can be achieved by randomization of data acquisition. The actual signal can be recovered via an efficient algorithm based on convex optimization methods. Compressed sensing has been applied to certain quantum tomography tasks. Standard compressed sensing has been directly used for efficient pseudothermal ghost imaging [20, 21]. Recently, a quadratic reduction in the total number of measurements for quantum tomography of a low rank density matrix has been demonstrated using a compressed sensing approach [22].

Here, we first describe how the Hamiltonian information is compressed into the experimental data. The output of a single measurement is related to the unknown signal (Hamiltonian parameters) through the relation (3). Suppose we try m different experimental configurations (i.e., m different pairs of $(M_j, |\psi_k\rangle)$). This yields a set of linear equations

$$\overrightarrow{p'} = \Phi\overrightarrow{h} \qquad (4)$$

where Φ is a $m \times d^2$ matrix with elements $\Phi_{jk,\alpha} = it/\sqrt{m}\left\langle\psi_k\right|[\Gamma_\alpha, M_j]\left|\psi_k\right\rangle$ (A factor $1/\sqrt{m}$ is included for simplifying the proofs, appendix C). In general m has to be

greater than or equal to d^2 in order to solve Eq. (4). A Hamiltonian estimation attempt with $m < d^2$ seems impossible as we face an underdetermined system of linear equations with an infinite number of solutions. However, any two s-sparse Hamiltonians h_1 and h_2 still can be distinguished via a properly designed experimental setting, if the measurement matrix Φ preserves the distance between h_1 and h_2 to a good approximation:

$$(1-\delta_s)||h_2-h_1||_{l_2}^2 \leq ||\Phi(h_2-h_1)||_{l_2}^2 \leq (1+\delta_s)||h_2-h_1||_{l_2}^2 \tag{5}$$

for a constant $\delta_s \in (0,1)$. A smaller δ_s ensures higher distinguishability of s-sparse Hamiltonians. The inequality relation (5) is termed a *restricted isometry property* (RIP) of the matrix Φ [23]. We now discuss how to construct a map Φ satisfying this inequality, and how small the value of m can be made.

The RIP (5) for a matrix Φ can be established by employing the measure concentration properties of random matrices. In each experiment the test state and the measurement observable can be drawn randomly from a set of configurations $\{M_j, |\psi_k\rangle\}$ realizable in the laboratory. The independent selection of $|\psi_k\rangle$ and M_j leads to a matrix Φ with independent rows but correlated elements $\Phi_{jk,\alpha}$ in each row. Thus the standard results from compressed sensing theory are not applicable here (appendix C).

In contrast, here we derive a concentration inequality for a matrix with independent rows and correlated columns as the backbone for the RIP of our quantum problem in appendix C. Using Hoeffding's inequality, we show that for any Hamiltonian h and a random matrix Φ with column only correlations, the random variable $||\Phi h||^2$ is concentrated around $||h||^2$ with a high probability, i.e. $\forall\ 0 < \delta < 1$

$$\text{Prob.}\{|||\Phi h||_{l_2}^2 - ||h||_{l_2}^2| \geq \delta||h||_{l_2}^2\} \leq 2e^{-mc_0(\delta+c_1)^2} \tag{6}$$

for some constants c_0 and c_1.

Using the above inequality, now we can show how an exponential reduction in the minimum number of the required configurations can be achieved for Hamiltonian estimation. The inequality (6) is defined for any h while the inequality in the definition of RIP, Eq.(5), is for any s-sparse h. As shown in Ref. [24], there is an inherent connection between these two inequalities. It is proved that any matrix Φ satisfying (6) has RIP with probability greater than $1 - 2\exp(-mc_0(\delta_{\frac{s}{2}} + c_1)^2 + s[\log(d^4/s) + \log(12e/\delta_{\frac{s}{2}})])$. In addition, whenever $m \geq c_2 s\log(d^4/s)$, for a sufficiently large constant c_2 one can find a constant $c_3 \geq 0$ such that the likelihood of the RIP to be satisfied converges exponentially fast to unity as $1 - 2\exp(-c_3 m)$.

The set of experimental configurations defined by Eq (4), and the concentration properties given by Eq (5) and (6) can be understood as encoding the information of a sparse Hamiltonian into a space with a lower dimension. Next we need to provide an efficient method for decoding in order to recover the original Hamiltonian. The decoder is simply the minimizer of the l_1 norm of the signal h. Implementing this decoder is a special convex optimization problem, which can be solved via fast classical algorithms, yet not stricktly scalable. Furthermore, the encoding/decoding scheme is robust to noisy data as $||p' - \Phi h||_{l_2} \leq \epsilon$ where ϵ is the noise threshold. Note that ϵ includes the error of linearization (see Eq.(2)) that is $\mathcal{O}(\sqrt{m}Kt^2)$. Denote h_0 as the true representation of the Hamiltonian. For a threshold η, $h_0(s)$ is an approximation to h_0 obtained by selecting the s elements of h_0 as those that are larger than ηh_{max} and setting the remaining elements to zero. Now we state our main result:

IV. ALGORITHM EFFICIENCY

If the measurement matrix $\Phi \in \mathbb{C}^{m\times d^4}$ is drawn randomly from a probability distribution that satisfies the concentration inequality in (5) with $\delta_s < \sqrt{2}-1$, then there exist constants $c_2, c_3, d_1, d_2 > 0$ such that the solution $h^\star$ to the convex optimization problem,

$$\begin{aligned} &\textit{minimize } ||h||_{l_1} \\ &\textit{subject to } ||p' - \Phi h||_{l_2} \leq \epsilon, \end{aligned} \tag{7}$$

satisfies,

$$||h^\star - h_0||_{l_2} \leq \frac{d_1}{\sqrt{s}}||h_0(s) - h_0||_{l_1} + d_2\epsilon \tag{8}$$

with probability $\geq 1 - 2e^{-mc_3}$ provided that,

$$m \geq c_2 s\log(d^4/s), \tag{9}$$

where the performance of a l_1 minimizer, Eq. (8), and the necessary bound $\delta_s < \sqrt{2}-1$ are derived by Candés in Ref. [25].

As an example, for a system consisting of n interacting qubits, the exponential number of parameters describing the dynamics, 2^{2n}, can be estimated with a linearly growing number of experiments $m \geq c_2 s(8\log(2)n - \log(s))$. The second term, $d_2\epsilon$, indicates that the algorithmic performance is bounded by the experimental uncertainties. Consequently, for fully sparse Hamiltonians and $\epsilon = 0$ the exact identification of an unknown Hamiltonian is achievable. The properties of the ensemble from which the states and measurement observables are chosen would determine the parameter δ_s and consequently the performance of the algorithm. The linear independency of the Φ matrix rows for a random set of local state preparations and observables can be guaranteed by a polynomial level of computational overhead before conducting the experiments.

A certification for the nearly sparsity assumption can be obtained from Eqs.(8) and (9) as follows: Suppose $h_m^\star$ is the algorithm's outcome for m configurations. The nearly sparsity assumption is certified on the fly during the experiment, if the estimation improvement $||h_{m+1}^\star - h_m^\star||$ converges to zero for a polynomially large total number of configurations.

V. PHYSICALLY NEARLY SPARSE HAMILTONIAN

Although physical systems at the fundamental level involve local two-body interactions, many-body Hamiltonians often

describe quantum dynamics in a particular representation or in well defined approximate limits. The strength of the non-local k-body terms typically is much smaller than the two-body terms with strength J and decreases with the number k. For a fixed sparsity threshold η, k_η is defined as the largest number k for which k-body terms have strength larger than ηJ. Then the number of the elements of a s-sparse approximation of a n-body Hamiltonian grows linearly as $\mathcal{O}(ng(k_\eta))$, where the $g(k_\eta)$ is determined by the geometry of the system.

A general class of many-body interactions arises when we change the basis for a bosonic or ferminoic system expressed by a (typically local) second-quantized Hamiltonian to a Pauli basis, e.g., via a Jordan-Wigner transformation. For fermionic systems the interactions are imposed physically from Coulomb's force and Pauli exclusion principle. The second-quantized Hamiltonian for these systems can be generally written as:

$$\hat{H} = \sum_{p,q} b_{pq} \hat{a}_p^+ \hat{a}_q + \sum_{p,q,r,s} b_{pqrs} \hat{a}_p^+ \hat{a}_q^+ \hat{a}_r \hat{a}_s, \qquad (10)$$

where the annihilation and creation operators ($\hat{a}_j$ and $\hat{a}_j^+$ respectively) satisfy the fermionic anti-commutation relations: $\{\hat{a}_i, \hat{a}_j^+\} = \delta_{ij}$ and $\{\hat{a}_i, \hat{a}_j\} = 0$ [26]. For example, in chemical systems the coefficients h_{pq} and h_{pqrs} can be evaluated using the Hartree-Fock procedure for N single-electron basis functions. The Jordan-Wigner transformation can then be used to map the fermionic creation and annihilation operators into a representation in terms of Pauli matrices $\hat{\sigma}^x$, $\hat{\sigma}^y, \hat{\sigma}^z$. This allows for a convenient implementation on a quantum computer, as was demonstrated recently for the efficient simulation of chemical energy of molecular systems [27]. An important example of a Coulomb based Hamiltonian is the spin-coupled interactions in quantum dots which has the following Pauli representation:

$$H = \sum_{i,j,k,\cdots} b_{i,j,k,\ldots} \sigma_A^i \otimes \sigma_B^j \otimes \sigma_C^k \cdots, \qquad (11)$$

where $A, B, C, \cdots$ indicate the location of the quantum dots, , σ^is are Pauli operators, and $b_{i,j,k,\ldots}$ generally represents a many-body spin interacting term. In practice, these Hamiltonians are highly sparse or almost sparse due to symmetry considerations associated with total angular momentum [28]. For example the Hamiltonian for the case of four quantum dots (A, B, C, D) takes the general form [28]:

$$H_{exchange} = J \sum_{A \le i < j \le D} \sigma_i.\sigma_j + J'[(\sigma_A.\sigma_B)(\sigma_C.\sigma_D) + (\sigma_A.\sigma_C)(\sigma_B.\sigma_D) + (\sigma_A.\sigma_D)(\sigma_B.\sigma_C)], \qquad (12)$$

Another class of effective many-body interactions often emerge in a perturbative and/or short time expansion of dynamics, such as effective three-body interactions between atoms in optical lattices [29] that we study in this work.

Next, we simulate the performance of our algorithm for estimation of such sparse many-body Hamiltonians in optical lattices [29] and quantum dots [28].

A. Three-body interactions in optical lattices

An optical lattice is a periodic potential formed from interference of counterpropagating laser beams where neutral atoms are typically cooled and trapped one per site. Consider four sites in two adjacent building blocks of a triangular optical lattice filled by two species of atoms [29]. The interaction between atoms is facilitated by the tunneling rate J between neighboring sites and collisional couplings U when two or more atoms occupy the same site. For each site an effective spin is defined by the presence of one type of atom as the up-state $\uparrow$ and the presence of the other type as the down-state $\downarrow$. Three-body interactions between atoms in a triangular optical lattice can be significant. The effective Hamiltonian for this system is studied in Ref. [29]. The on-site collisional interaction U, and tunneling rates $J = J^\uparrow = 2J^\downarrow$ are taken to be the same in all sites, also $U = U_{\uparrow\uparrow} = U_{\downarrow\downarrow} = 2.12U_{\uparrow\downarrow} = 10kHz$. The effective Hamiltonian of the 4-spin system is

$$H_{opt-latt} = \sum_{j,\alpha=x,y,z} b_1^\alpha \sigma_j^\alpha \sigma_{j+1}^\alpha + b_2^\alpha \sigma_j^\alpha \sigma_{j+1}^\alpha \sigma_{j+2}^\alpha \qquad (13)$$

where $\{b_1^\alpha, b_2^\alpha\}$ are functions of $\{J, U\}$ and their explicit forms are given in appendix D. The ratio $\eta = |J/U|$ quantifies the sparsity level. For a fixed value of U, a smaller J leads to weaker three-body interactions and therefore a higher level of sparsity. As expected, this enhances the algorithm performance.

We assume that the system can be initialized in a random product state $|\psi_k\rangle = |\psi_k^1\rangle \otimes ... \otimes |\psi_k^4\rangle$, where $|\psi_k^i\rangle$ are drawn from the Fubini-Study metric induced distribution. The required observables for the algorithm are uniformly selected from single qubit Pauli operators $\{\sigma_i^x, \sigma_i^y, \sigma_i^z\}$. This choice of states and observables allows for $\delta_s \approx 0.37 < \sqrt{2} - 1$. Let us denote the extracted Hamiltonian and the true Hamiltonian by H^* and H_{true}, respectively. Here, the performance of the algorithm is defined by the relative error $1 - ||H^* - H_{true}||_{fro}/||H_{true}||_{fro}$. The results for different number of configurations are depicted in Fig. (1), for various values of J. As evident in Fig.(1), performance accuracy of above 94% can be obtained with only 80 settings significantly smaller than approximately 6×10^4 configurations required in QPT.

The robustness of this scheme was also investigated for 10% random error in simulated experimental data leading to about a 5% reduction in the overall performance.

B. Four-body interactions in quantum dots

Another important class of effective many-body Hamiltonians can be obtained for electrons in quantum dots coupled through an isotropic (Heisenberg) or anisotropic exchange interaction. For example the Hamiltonian for the case of four quantum dots (A, B, C, D) takes the general form Eq. (12). The first term in the summation is a two-body Heisenberg exchange interaction and the last three terms are four-body spin interactions. In certain regimes, the ratio $|J'/J|$ can reach up

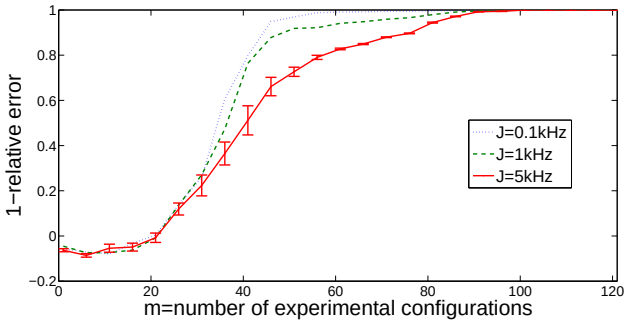


FIG. 1: The Hamiltonian estimation average performance is illustrated for a system of four adjacent sites in an optical lattice for different tunneling rates, J, and collisional coupling $U = 10kHz$. The error bars demonstrate the standard deviation of the performance due to the random and independent selection of m configurations (shown only for $J = 5kHz$). Performance accuracy of above 90% with only 60 settings is achievable for $J = 1kHz$, which is significantly smaller than about 6×10^4 required experimental configurations in QPT.

to 16%. The amplitude of $\eta = |J'/J|$ determines the sparsity level of the Hamiltonian.

Here we use an efficient modification of signal recovery referred as "reweighted l_1-minimization" which is described in appendix E. The performance of this algorithm is demonstrated in Fig. (2) that shows a significant reduction of the required number of settings in contrast to the standard QPT.

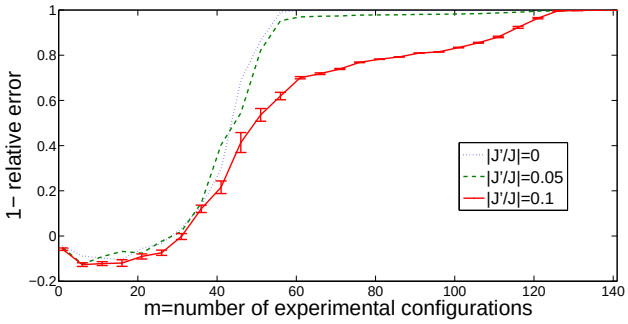


FIG. 2: Estimation of the exchange interaction Hamiltonian for four electrons in quantum dots. The average performance of the procedure is illustrated for different values of $|J'/J|$ with 50 iterations of the l_1-reweighted minimization. The standard deviations are shown only for $|J'/J| = 0.1$ It is demonstrated that only 60 different configurations are sufficient for estimating the unknown Hamiltonian with an accuracy above 95% for $|J'/J| = 0.05$, instead of about 6×10^4 required settings via QPT.

VI. V. CHARACTERIZATION OF HAMILTONIAN FINE STRUCTURES AND SYSTEM-BATH INTERACTIONS

A. Hamiltonian fine estimation

In many systems a primary model of the interactions is often known through physical and/or engineering considerations. Starting with such an initial model we seek to improve our knowledge about the Hamiltonian by random measurements. Let's assume the initial guess about the Hamiltonian H_0 is close to the true form H_{true} that is $||\Delta = H_{true} - H_0|| \ll ||H_{true}||$. Therefore for a perturbative treatment we demand $t||\Delta|| \ll 1$, which is a much weaker requirement compared to $t||H_{true}|| \ll 1$. We can approximate Eq. (1) in the paper to find

$$p_{jk} \approx \langle\psi_k| M_j^0 |\psi_k\rangle + i \langle\psi_k| [\int_0^t e^{isH_0}\Delta e^{-isH_0} ds, M_j^0] |\psi_k\rangle , \quad (14)$$

where $M_j^0 = e^{itH_0} M_j e^{-itH_0}$ [31]. This equation is linear in Δ, consequently, in a similar fashion as above, the compressed sensing analysis can be applied for efficient estimation of the fine structure of Hamiltonians.

B. Characterizing system-bath interactions

The identification of a decoherence process is a vital task for quantum engineering. In contrast to the usual approach of describing dynamics of an open quantum system by a Kraus map or a reduce master equation, here we use a microscopic Hamiltonian picture to efficiently estimate the system-bath coupling terms generating the overall decoherence process. However since we consider a full dynamics of the system and bath, this method can be applied to a finite size environment such as a spin bath, or a surrogate Hamiltonian modeling of a infinite bath. In the latter case a harmonic bath of oscillators is approximated by a finite spin bath [32].

Consider an open quantum system with a total Hamiltonian:

$$H = H_S \otimes I_B + I_S \otimes H_B + H_{SB} \quad (15)$$

and

$$H_{SB} = \sum_{p,q} \lambda_{p,q} S_p \otimes B_q \quad (16)$$

where H_S (H_B) denotes the system (bath) free Hamiltonian and H_{SB} is the system-bath interaction with coupling strengths $\{\lambda_{p,q}\}$, and a complete operator basis of the system and bath being $\{S_p\}$ and $\{B_q\}$, respectively.

We develop a formalism to estimate $\lambda_{p,q}$ parameters in the weak system-bath coupling regime and with the sparsity assumption that a few number of $\lambda_{p,q}$ have a significant value.

The Liouvilian dynamical equation is

$$\frac{d}{dt}\rho_{SB}(t) = (\mathcal{L}_0 + \sum_{pq} \lambda_{pq}\mathcal{L}_{pq})[\rho_{SB}(t)] \quad (17)$$

where $\mathcal{L}_0[.] = -i[H_S \otimes I_B + I_S \otimes H_B, .]$ and $\mathcal{L}_{pq}[.] = -i[S_p \otimes B_q, .]$. In the regime of weak coupling to a finite bath, $||H_{SB}|| \ll \min\{||H_S||, ||H_B||\}$, the Liouvillan equation (17) can be solved perturbatively if time t satisfies $t||H_{SB}|| \ll 1$. For an initial system density state ρ_k, using the matrix identity given in Ref. [31] we find the measurement outcomes as

$$p_{jk} \approx tr(\rho_k M_j) + \sum_{pq} \lambda_{pq} tr([\int_0^t ds e^{(t-s)\mathcal{L}_0} \mathcal{L}_{pq} e^{s\mathcal{L}_0}[\rho_k], M_j]) \quad (18)$$

where M_j is a system only observable. This affine function between the outcomes p_{jk} and coupling parameters $\{\lambda_{pq}\}$ is similar to Eq.(2) in the paper for Hamiltonian estimation. Consequently, the compressed sensing algorithm can be employed for computing $\{\lambda_{pq}\}$s.

VII. OUTLOOK

We have introduced an efficient and robust experimental procedure for the identification of nearly sparse Hamiltonians using only separable (local) random state preparations and measurements. There are a number of future directions and open problems associated with this work. It is not known how the performance of the algorithm depends on the distribution of the ensemble from which the states and measurement observables are drawn. Also, a general closed-loop learning approach for updating the knowledge of sparsity basis of an arbitrary Hamiltonian is an interesting open problem that will be of importance for generic compressed system identification. The presented method for Hamiltonian estimation is promising for drastic reduction in the number of experimental configurations. However the classical resources for post-processing is not scalable. A fully scalable Hamiltonian estimation method might be achievable via a hybrid of compressed sensing and DMRG (Density-Matrix Renormalization Group) methods [33]. A compressed tomography method can also be developed for nearly sparse quantum processes [34].

VIII. ACKNOWLEDGEMENT

We thank NSERC and Center for Extreme Quantum Information Theory (MM), and DARPA Grant FA9550-09-1-0710 (RLK, HR) for funding.

Appendix A: vectors and operator norm

In this paper we use the following different norms:
For a vector x,

$$||x||_{l_2} = \sqrt{x^\dagger x}, ||x||_{l_1} = \sum_i |x_i|. \quad (A1)$$

For a matrix A,

$$||A||_{spec} = \sqrt{\lambda_{max}(A^\dagger A)} \quad (A2)$$

where λ_{max} means largest eigenvalue.

$$||A||_{fro} = \sqrt{trace(A^\dagger A)} \quad (A3)$$

Appendix B: Analysis of the short time approximation

The short time monitoring of the system's dynamics requires a prior knowledge of the dynamical time scales. In the solid-state quantum devices, in particular in the context of quantum control and quantum information-processing, the time-scale of single qubit rotations is typically on the order of 1-10 ns. The switching time for exchange interactions varies among different solid-state systems. For superconducting phase qubit the duration of a swap gate is about 10 ns [35]. For electron-spin qubits in quantum dots and in donor atoms (Heisenberg models) [36–38], and also for quantum dots in cavities (anisotropic exchange interactions) [39] the coupling time is between 10-100ps, while for exciton-coupled quantum dots (XY model) and Forster energy transfer in multichromophoric complexes the relevant time scale is in the order of 1ps. Next we rigorously derive bound on the evolution time t that guarantees the validity of the short time approximation.

For an input state $|\psi_k\rangle$, the expectation value of an observable M_j is

$$p_{jk} = \langle\psi_k(t)| M_j |\psi_k(t)\rangle = \langle\psi_k| e^{iHt} M_j e^{-iHt} |\psi_k\rangle \quad (B1)$$

Considering the expansion of the propagator $e^{-iHt} = I - itH - \frac{1}{2}t^2H^2 + ...$, we find

$$p_{jk} = \langle\psi_k| M_j |\psi_k\rangle + it \langle\psi_k| [H, M_j] |\psi_k\rangle - \frac{t^2}{2} \langle\psi_k| [H, [H, M_j]] |\psi_k\rangle + ... \quad (B2)$$

Therefore, for the linearization assumption, it is sufficient to have for the l'th term

$$t^l \min_j \langle\psi_k| \overbrace{[H, [H, [..., M_j]]]}^{l \text{ times}} |\psi_k\rangle \leq t^l \min_j ||[H, [H, [..., M_j]]]||_{spec} \ll 1, \forall l. \quad (B3)$$

A tighter bound can be found for operators $\{M_j\}$ from a POVM as

$$||[H, [H, [..., M_j]]]||_{spec} \leq 2^l ||H||_{spec}^l \quad (B4)$$

To derive this we use

$$||[A, B]||_{spec} \leq ||AB||_{spec} + ||BA||_{spec} \leq 2||A||_{spec}||B||_{spec} \quad (B5)$$

and $||A||_{spec}^2 = ||AA^\dagger||_{spec}$.

This gives a single bound sufficient for linearization: $t \ll \frac{1}{2}||H||_{spec}^{-1}$.

Appendix C: RIP from a concentration inequality

In this work, we generalize the standard compressed sensing algorithm such that the necessity for independent randomness in all elements of the measurement matrix, ϕ, can be avoided. A common approach to establish RIP ([24]) for a matrix Φ is by introducing randomness in the elements of this matrix. This approach benefits from measure concentration properties of random matrices. In classical signal processing each element $\Phi_{jk,\alpha}$ can be independently selected from a random distribution such as Gaussian or Bernoulli. Whereas in the Hamiltonian estimation formulation (Eq. (4) in the paper) there is no freedom for independent selection of the Φ matrix elements.

Here we prove the concentration inequality that we employed for establishing the restricted isometry property.

Though Φ is a random matrix, because it is constructed from quantum states and observables of a finite dimensional system, it is bounded. Thus we are able to apply *Hoeffding's concentration inequality:* If $v_1, ..., v_m$ are independent bounded random variables such that Prob.$\{v_i \in [a_i, b_i]\} = 1$, then for $S = \sum_i v_i$,

$$\begin{aligned} \text{Prob.}\{S - \mathbf{E}(S) \geq t\} &\leq e^{-2t^2/\sum_i (b_i - a_i)^2} \\ \text{Prob.}\{S - \mathbf{E}(S) \leq -t\} &\leq e^{-2t^2/\sum_i (b_i - a_i)^2} \end{aligned} \tag{C1}$$

for any $t > 0$. (Here $\mathbf{E}$ denotes the expectation value.) Set $v_i = |\phi_i^\dagger x|^2$ for a row ϕ_i. Then with $S = \sum_i v_i = ||\Phi x||_{l_2}^2$, we get $\forall x$,

$$\begin{aligned} v_i = x^\dagger(\phi_i \phi_i^\dagger)x \in (1/m)[w_l, w_u]||x||_{l_2}^2 \\ \mathbf{E}(S) = \mathbf{E}||\Phi x||_{l_2}^2 \in [f, g]||x||_{l_2}^2 \end{aligned} \tag{C2}$$

for constants w_l, w_u, f, g. Note that f and g are the min and max singular values of $\mathbf{E}(\Phi^\dagger \Phi)$. From (C2) we find $\forall t_+, t_- > 0$ and $\forall x$,

$$\begin{aligned} \text{Prob.}\{S - g||x||_{l_2}^2 \geq t_+\} &\leq \text{Prob.}\{S - \mathbf{E}(S) \geq t_+\} \\ \text{Prob.}\{S - f||x||_{l_2}^2 \leq -t_-\} &\leq \text{Prob.}\{S - \mathbf{E}(S) \leq -t_-\} \end{aligned}$$

These together with (C1) and (C2), and the choice of $t_+ = (\delta + 1 - g)||x||_{l_2}^2$ and $t_- = (f - 1 + \delta)||x||_{l_2}^2$ yields

$$\text{Prob.}\{||\Phi x||_{l_2}^2 - ||x||_{l_2}^2| \geq \delta ||x||_{l_2}^2\} \leq 2e^{\frac{-2m(\delta+\epsilon)^2}{(w_u - w_l)^2}} \tag{C3}$$

with $\epsilon = \min\{1 - g, f - 1\}$. To ensure that $t_+, t_- > 0$, we need $1 - \delta < f \leq g < 1 + \delta$. Since the observable M can be scaled by any real number, a sufficient condition is $g/f < (1 + \delta)/(1 - \delta)$. For the simulations in this paper, this ratio becomes 2.176.

Appendix D: 4-sites optical lattice Hamiltonian

Let us consider four sites in two adjacent building blocks of a triangular optical lattice filled by two species of atoms, $\uparrow$ and $\downarrow$. Atoms interact by tunneling between neighboring sites, $J^\uparrow$ and $J^\downarrow$, and through collisional couplings in the same site, U. The Hamiltonian for such system can be written as [29]:

$$\begin{aligned} H_{opt-latt} = \sum_j (0.03\frac{J^{\uparrow 2} + J^{\downarrow 2}}{U} - 0.27\frac{J^{\uparrow 3} + J^{\downarrow 3}}{U^2})\sigma_j^z \sigma_{j+1}^z \\ -(\frac{2.1(J^\uparrow + J^\downarrow)J^\uparrow J^\downarrow}{U^2} + \frac{J^\uparrow J^\downarrow}{U})(\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y) \\ + \sum_j 0.14\frac{J^{\uparrow 3} - J^{\downarrow 3}}{U^2}\sigma_j^z \sigma_{j+1}^z \sigma_{j+2}^z \\ -0.6\frac{J^\uparrow J^\downarrow (J^\uparrow - J^\downarrow)}{U^2}(\sigma_j^x \sigma_{j+1}^z \sigma_{j+2}^x + \sigma_j^y \sigma_{j+1}^z \sigma_{j+2}^y), \end{aligned} \tag{D1}$$

where $\sigma_j^{x,y,z}$ are Pauli operators.

Appendix E: Reweighted l_1-minimization

In order to simulate our alogrithm performance for estimating the above Hamiltonian we use an iterative algorithm that outperforms the standard l_1 norm minimization [30]. This procedure entails initializing a weight matrix $W = I_{d^2}$ and a weight factor $\sigma > 0$, and repeating the following steps until convergence is reached:

1. Solve for h, minimize $||Wh||_{l_1}$
subject to $||p' - \Phi h||_{l_2} \leq \epsilon$.
2. Update weights
$W = diag(1/(|h_1| + \sigma), ..., 1/(|h_{d^2}| + \sigma))$. (E1)

where $h = vec(h_i)$ is the Hamiltonian vectorized form. Φ is the measurement matrix and p' is the experimental data with a noise threshold ϵ.

[1] I.L. Chuang and M.A. Nielsen, J. Mod. Opt. 44, 2455. (1997).
[2] D. Leung Ph.D. Thesis (2000).
[3] D. W. Leung J. Math. Phys. **4**, 528 (2003).
[4] J. Emerson *et al.* Science **302**, 2098 (2003).
[5] G. M. D'ariano, P. L. Presti Phys. Rev. Lett. **91**, 047902 (2003).
[6] M. Mohseni and D. A. Lidar, Phys. Rev. Lett. **97**, 170501(2006).
[7] J. Emerson *et al.* Science **317**, 1893 (2007).
[8] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
[9] J. M. Geremia and H. Rabitz, Phys. Rev. Lett. **89**, 263902 (2002).

[10] N. Boulant, *et al.*, Phys. Rev. A **67**, 042322 (2003).
[11] R. L. Kosut, *et al.*, quant-ph/0411093 (2004).
[12] K. C. Young, *et al.*, Phys. Rev. A **79**, 062301 (2009).
[13] J. H. Cole, *et al.*, Phys. Rev. A **71**, 062312 (2005).
[14] S. J. Devitt, *et al.*, Phys. Rev. A **73**, 052317 (2006).
[15] S. G. Schirmer and D. K. L. Oi, Phys. Rev. A **80**, 022333 (2009).
[16] B. Levi, *et al.*, Phys. Rev. A **75**, 022314 (2007).
[17] M. Mohseni and A. Rezakhani, Phys. Rev. A **80**, 010101 (2009).
[18] C. Di Franco *et al.* Phys. Rev. Lett. **102**, 187203 (2009).
[19] E. J. Candés and M. B. Wakin, IEEE Sig. Proc. Mag. **Mar.**, 21 (2008).
[20] O. Katza, Y. Bromberg, and Y. Silberberg Appl. Phys. Lett., **95**, 131110 (2009).
[21] W. Gong and S. Han arXiv:0910.4823v1, (2009).
[22] D. Gross *et al.*, arXiv:0909.3304 (2009).
[23] The common definition of RIP is $(1-\delta_{s/2})||h||_{l_2}^2 \leq ||\Phi h||_{l_2}^2 \leq (1+\delta_{s/2})||h||_{l_2}^2$ for a s-sparse h [19] which is equivalent to the definition (5) in the paper.
[24] R. Baraniuk *et al.*, Constr. Approx. **28**, 253 (2008).
[25] E. J. Candés, Compte Rendus de l'Academie des Sciences, Paris, Serie I, **346**, 589 (2008).
[26] G. D. Mahan, *Many-particle physics* (Springer, 2000).
[27] B. P. Lanyon, *et al.*, Nature Chemistry **2**, 106 (2009).
[28] A. Mizel and D. Lidar, Phys. Rev. B **70**, 115310 (2004).
[29] J. K. Pachos and E. Rico, Phys. Rev. A **70**, 053620 (2004).
[30] E.J. Candes,*et al.*, J. Fourier Anal. Appl. **14**, 877-905 (2007).
[31] Equations (14) and (18) are direct applications of the matrix identity $e^{(A+B)t} = e^{At} + \int_0^t e^{A(t-s)} B e^{(A+B)s} ds$.
[32] D. Gelman, *et al.*, J. Chem. Phys. **121**, 661 (2004).
[33] M. Cramer, M. B. Plenio, arXiv:1002.3780 (2010).
[34] A. Shabani, R. L. Kosut, H. Rabitz, arXiv:0910.5498 (2009).
[35] R. C. Bialczak, *et al.*, arXiv:0910.1118 (2009).
[36] D. Loss and D. P. DiVincenzo, Phys. Rev. A **57**, 120 (1998).
[37] B.E. Kane, Nature **393**, 133 (1998).
[38] R. Vrijen, *et al.*, Phys. Rev. A **62**, 012306 (2000).
[39] A. Imamoglu, *et al.*, Phys. Rev. A **83**, 4204 (1999).

References

[1] G. R. Fleming and M. A. Ratner. Grand challenges in basic energy sciences. *Physics Today*, 61-7:28–33, July 2008.

[2] Arda quantum computing roadmap. available at http://qist.lanl.gov, 2004.

[3] P. Dowling and G. J. Milburn. Quantum technology: The second quantum revolution. *Proceedings of the Royal Society of London*, A 361:1655, 2003.

[4] M.Q. Phan and H. Rabitz. Learning control of quantum-mechanical systems by laboratory identification of effective input-output maps. *Chem. Phys.*, 217:389–400, 1997.

[5] J. M. Geremia and H. Rabitz. Teaching lasers to optimally identify molecular hamiltonians. *Phys. Rev. Lett.*, 89 263902, 2002.

[6] M. Reimpell and R. F. Werner. Iterative optimization of quantum error correcting codes. *Phys. Rev. Lett.*, 94:080501, 2005.

[7] A. S. Fletcher, P. W. Shor, and M. Z. Win. Optimum quantum error recovery using semidefinite programming. *Phys. Rev. A*, 75:012338, 2007. quant-ph/0606035.

[8] R. L. Kosut, A. Shabani, and D. A. Lidar. Robust quantum error correction via convex optimization. *Phys. Rev. Lett.*, 100:020502, 2008. (arXiv:quant-ph/0703274).

[9] H. Rabitz. Making molecules dance: Optimal control of molecular motion. In A.D. Bandrauk, editor, *Atomic and Molecular Processes with Short Intense Pulses*. Plenum Publishing Corporation, 1988.

[10] H. Rabitz, T.S. Ho, M. Hsieh, R. Kosut, and M. Demiralp. Topology of optimally controlled quantum mechanical transition probability landscapes. *Phys. Rev. A*, 74(012721), 2006.

[11] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4), April 2006.

[12] E. J. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, Aug. 2006.

[13] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, (6):1182–1195, December 2007.

[14] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008.

[15] R. L. Kosut. Quantum process tomography via ℓ_1-norm minimization. *arXiv:0812.4323v1[quant-ph]*, 2008.

[16] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*. Addison-Wesley, 1963-1965.

[17] K. C. Young, M. Sarovar, R. L. Kosut, and K. B. Whaley. Optimal quantum multi-parameter estimation as applied to dipole- and exchange-coupled qubits. *Physical Review A*, 79:062301, 2009.

[18] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, January 2006.

[19] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam, 1982.

[20] S. Braunstein and C. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439, 1994.

[21] M. Sarovar and G. J. Milburn. Optimal estimation of one parameter quantum channels. *J. Phys. A: Math. Gen.*, 39:8487–8505, 2006.

[22] S. Boixo, S.T. Flammia, C.M. Caves, and JM Geremia. Generalized Limits for Single-Parameter Quantum Estimation. *Phys. Rev. Lett.*, 98:090401, 2007.

[23] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde. Entanglement-free heisenberg-limited phase estimation. *Nature*, 450:393–396, Nov. 2007.

[24] C. Brif and A. Mann. Nonclassical interferometry with intelligent light. *Phys. Rev. A*, 54:4505, 1996.

[25] U. Dorner, R. Demkowicz-Dobrzanski, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley. Optimal quantum phase estimation. *quant-ph/0807.3659*, 2008.

[26] R. L. Kosut. Quantum metrology subject to instrumentation constraints. *quant-ph/arXiv:0803.4284*, March 2008.

[27] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.